

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2003, 06:21:26 ; Search time 6392 Seconds  
(without alignments)  
17292.312 Million cell updates/sec

Title: US-09-763-985A-1  
Perfect score: 3798  
Sequence: 1 ccacgcgtccgagtcgagc.....tcaaaaaaaaaaaaaaaaaa 3798

Scoring table: IDENTITY\_NUC  
Gap 10.0, Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues  
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.om.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.mu.\*
- 20: em.om.\*
- 21: em.or.\*
- 22: em.ov.\*
- 23: em.pat.\*
- 24: em.ph.\*
- 25: em.pl.\*
- 26: em.ro.\*
- 27: em.sts.\*
- 28: em.un.\*
- 29: em.vi.\*
- 30: em.htg\_hum.\*
- 31: em.htg\_inv.\*
- 32: em.htg\_other.\*
- 33: em.htg\_mus.\*
- 34: em.htg\_pln.\*
- 35: em.htg\_rtd.\*
- 36: em.htg\_mam.\*
- 37: em.htg\_vrt.\*
- 38: em.sy.\*
- 39: em.htgo\_hum.\*
- 40: em.htgo\_mus.\*
- 41: em.htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	3771.2	99.3	3788	9	AB020880	AB020880 Homo sapi
2	3656.8	96.3	3660	9	D63879	D63879 Homo sapi
3	3611.2	95.1	3644	9	AF387506	AF387506 Human mRNA
4	2886	76.0	3537	6	AX268309	AX268309 Sequence
5	2154.2	56.7	3586	10	AF172722	AF172722 Mus muscu
6	1056.4	27.8	171523	9	AC010206	AC010206 Homo sapi
7	1056.4	27.8	171523	9	AC008119	AC008119 Homo sapi
8	769.4	20.3	771	11	G26688	G26688 human STS S
9	338.2	8.9	86096	2	AC122224_3	Continuation (4 of
10	338.2	8.9	319737	2	AC126267	AC126267 Mus muscu
11	306.2	8.1	94682	2	AC128917	AC128917 Rattus no
12	233.4	6.1	324	6	AX312274	AX312274 Sequence
13	194	5.1	314	11	G62164	G62164 EST1135 Poe
14	186.8	4.9	218	9	HS53G8F	Z55627 H.sapiens C
15	162	4.3	369	11	G54201	G54201 B13G14/T7 H
16	157.4	4.1	180819	2	AC126857	AC126857 Rattus no
17	150.2	4.0	90963	2	AC079626	AC079626 Mus muscu
18	133.4	3.5	207933	2	AC128551	AC128551 Rattus no
19	123.2	3.2	180819	2	AC126857	AC126857 Rattus no
20	112.4	3.0	13810	2	AL845326	AL845326 Danio rer
21	97.8	2.6	18858	2	AC079867	AC079867 Mus muscu
22	95.6	2.5	7218	6	I66494	I66494 Sequence 14
23	90.6	2.4	295	9	HS183C8F	Z57540 H.sapiens C
24	86	2.3	94682	2	AC128917	AC128917 Rattus no
25	75	2.0	2628	8	AY075670	AY075670 Arabidops
26	67	1.8	197346	2	AL672124	AL672124 Mus muscu
27	66.6	1.8	195890	10	AC098726	AC098726 Mus muscu
28	66.6	1.8	207997	2	AL845462	AL845462 Mus muscu
29	64.2	1.7	215219	2	AC118225	AC118225 Mus muscu
30	64.2	1.7	234817	10	AL663048	AL663048 Mouse DNA
31	63.4	1.7	179145	9	AC009785	AC009785 Homo sapi
32	63.4	1.7	194334	2	AC101527	AC101527 Mus muscu
33	63.4	1.7	198228	2	AC102097	AC102097 Mus muscu
34	63.2	1.7	168528	9	AL355861	AL355861 Human DNA
35	63	1.7	177269	2	AC108444	AC108444 Mus muscu
36	62.8	1.7	195481	2	AC113303	AC113303 Mus muscu
37	62.6	1.6	160653	10	AC122186	AC122186 Mus muscu
38	62.4	1.6	228664	2	AL844566	AL844566 Mus muscu
39	62.2	1.6	31930	2	AC073495	AC073495 Mus muscu
40	62	1.6	3127	3	DM4FRNP	Y15251 Drosophila
41	62	1.6	3358	3	AY061395	AY061395 Drosophila
42	62	1.6	157320	2	AC118889	AC118889 Rattus no
43	62	1.6	172665	10	AL645740	AL645740 Mouse DNA
44	62	1.6	203769	2	AL603745	AL603745 Mus muscu
45	62	1.6	208070	2	AL844169	AL844169 Mus muscu

ALIGNMENTS

RESULT 1  
AB020880  
LOCUS  
DEFINITION Homo sapiens mRNA for squamous cell carcinoma antigen SART-3,  
complete cds.  
3788 bp mRNA linear PRI 03-JUN-1999  
ACCESSION AB020880  
VERSION AB020880.1 GI:4996281  
KEYWORDS squamous cell carcinoma antigen SART-3.  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 3788)  
AUTHORS Itoh,K., Yang,D., Sasatomi,T., Nakao,M., Shichijo,S., Takasu,H.,  
Matsumoto,H., Mori,K. and Yamana,H.

TITLE SART-3 (Squamous cell carcinoma antigen recognized by T cells)  
JOURNAL Published Only in DataBase (1999)  
REFERENCE 2 (bases 1 to 3788)  
AUTHORS Itoh, K., Yang, D., Sasatomi, T., Nakao, M., Shichiho, S., Takasu, H.,  
Matsumoto, H., Mori, K. and Yamana, H.  
TITLE Direct Submission  
JOURNAL  
of Immunology; 67 Asahi-machi, Kurume University, Department  
(E-mail: kyogedmed.kurume-u.ac.jp, Tel: 81-942-35-3311,  
Fax: 81-942-31-7699)

FEATURES Location/Qualifiers  
source  
1..3788  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/cell\_line="KE4"  
/cell\_type="squamous cell carcinoma"  
20..2911  
/note="squamous cell carcinoma antigen recognized by T  
cells"  
/codon\_start=1  
/product="squamous cell carcinoma antigen SART-3"  
/protein\_id="BAA78384.1"  
/db\_xref="GI:4996282"

CDS  
1..3788  
/translation="MATAEETASPEAESKAPKADGEEDVEKAAARTRRKVLRAVA  
AATYKTMGPWQOQEGVSESDGEYMASSAPGEYENYDEEERKNOLEIRLE  
EQLSINRYNCHVDLIRLLRLEGELTKYRMARQKMSIFPLTELWLEWHLDEISMA  
QGLDRHYVDLFEKAVKDYICPNWLEYQYVSGVIGOKGLEKVRVFERALSVG  
LHMTGLWEDVREFSAIYVAAKLEKYSFLRROLAIPLYDMEATPAEYEWSEDP  
IPESVIONKALQLEKYPVEEALLOAEPLAEOYAYIDFENKIGDPARILQIFE  
RALVENCPLDWIRYSQYLDKQLKVDLIVSHNRAIRNCPTWALWRSYLLAMERH  
GVHQVISTFEKALNAGFIQATVYIWOAYDLRLRRVDFKQSSKELEELRAFT  
RALEYLUKVEEFERNAGDPSVIMONWARIELRNNMOKARELWDSIMTRGNKAY  
NMWLEYNLERAGHTQHCRLHRAVOCTSDYPEHVCVELLIMTEGSLIEDWDIAV  
QKTEYLRNVEQKRAEKAALVOOEKAEORRAEKKALKKKKIRGPEKRG  
ADDDKENGDDDEEOPSKRRVENSIPAGETONVEVAGPAGCAADVPEPSKQK  
EKAALRDPKPLVHDSKDSITVFNLSNIPYMQBPDTKRLPLFACRGVVOIRPFS  
NRGDFGYCIVFEKESALQALEMDRKSVEGRPMFVSPVDKSNPDKPFVRYSTSL  
EYHKLFISGLPFSCKEELEICKAHGTVDLRLVNRAGKPKGLAYVEYENESQSQ  
AVMMQDMGTIKENIKVAISNPQRKVPKPKTRKAPGGPMLLPOTYGARKGRTQLS  
LLRALRPSAAPQAENGPAAPAAVAAPAAEAPAKMSNADFALFLRK"  
3788  
polyA\_site  
/note="18 A nucleotides"

BASE COUNT 1020 a 846 c 1081 g 841 t  
ORIGIN

Query Match 99.3%; Score 3771.2; DB 9; Length 3788;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 3773; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 GGTCCGATGGCGACTGCGGCGGAAACCTCGGCTTCAGAACCCGAGGCTGAGTCCAAAGC 64  
DB 13 GCGCAAGATGGCGACTGCGGCGGAAACCTCGGCTTCAGAACCCGAGGCTGAGTCCAAAGC 72  
QY 65 TGGGCCCAAGGCTGACGAGAGAGGATGAGGTAAAGCGGCTAGGACAAAGGAAAGCT 124  
DB 73 TGGGCCCAAGGCTGACGAGAGAGGATGAGGTAAAGCGGCTAGGACAAAGGAAAGCT 132  
QY 125 GTTATCGCGGCTGTGCGCGCTCGCACATACAGACCATGGGCGCAGCGTGGGATCAGCA 184  
DB 133 GTTATCGCGGCTGTGCGCGCTCGCACATACAGACCATGGGCGCAGCGTGGGATCAGCA 192  
QY 185 GGAGGAAGCGGTGACGAGAGGATGGGATGAGTACGCCATCGGCTTCCTCGCGGAGAG 244  
DB 193 GGAGGAAGCGGTGACGAGAGGATGGGATGAGTACGCCATCGGCTTCCTCGCGGAGAG 252  
QY 245 CTCCTCCCGGGAGTACGAGTGGGAATATCAGCAAGAGGAGGAGAAACCAAGCTGGAGT 304  
DB 253 CTCCTCCCGGGAGTACGAGTGGGAATATCAGCAAGAGGAGGAGGAGAAACCAAGCTGGAGT 312  
QY 305 TGAGAGACTGGAGGAGCACTTGTCTATCAACGTCATGACTACAACTGCCATGTGGACTT 364  
DB 313 TGAGAGACTGGAGGAGCACTTGTCTATCAACGTCATGACTACAACTGCCATGTGGACTT 372

QY 365 GATCAGACTGCTCAGGCTGGAAGGGAGGCTTACCAAGGTGAGGATGCGCCGCCAGAAGAT 424  
DB 373 GATCAGACTGCTCAGGCTGGAAGGGAGGCTTACCAAGGTGAGGATGCGCCGCCAGAAGAT 432  
QY 425 GAGTGAATCTTTCCCTTCTGACTGAAGAGCTCTGGCTGGAGTGGCTGCATGACGAGATCAG 484  
DB 433 GAGTGAATCTTTCCCTTCTGACTGAAGAGCTCTGGCTGGAGTGGCTGCATGACGAGATCAG 492  
QY 485 CATGGCCCAAGGATGGCTTGACAGAGAGCAGCGTATGACCTCTTTTGAGAAAGCCCTGAA 544  
DB 493 CATGGCCCAAGGATGGCTTGACAGAGAGCAGCGTATGACCTCTTTTGAGAAAGCCCTGAA 552  
QY 545 GGATTACATTTGTCTTACATTTGGCTAGAGTATGGCCAGTACTACTAGTTGGTGGGATGG 604  
DB 553 GGATTACATTTGTCTTACATTTGGCTAGAGTATGGCCAGTACTACTAGTTGGTGGGATGG 612  
QY 605 TCAGAAAGTGGCTTGAGAAAGTTCGCTCCGCTGTTGAAAGGGCTCTCTCGTCTGTGG 664  
DB 613 TCAGAAAGTGGCTTGAGAAAGTTCGCTCCGCTGTTGAAAGGGCTCTCTCGTCTGTGG 672  
QY 665 TTTACATATGACCAAGGACTCGCCCTCTGGGAGGCTTACCGAGAGCTTTGAAAAGTGGAT 724  
DB 673 TTTACATATGACCAAGGACTCGCCCTCTGGGAGGCTTACCGAGAGCTTTGAAAAGTGGAT 732  
QY 725 TGTGGAAGTGTCTCGGCTTGAGAAAGTCCACAGTCTTTTCCGCGCAGAGTGTGGCGATCC 784  
DB 733 TGTGGAAGTGTCTCGGCTTGAGAAAGTCCACAGTCTTTTCCGCGCAGAGTGTGGCGATCC 792  
QY 785 ACTCTATGATATGGAGGCCACATTTGCAGAGTATGAAGATGGTCTCAGAGACCCCAATACC 844  
DB 793 ACTCTATGATATGGAGGCCACATTTGCAGAGTATGAAGATGGTCTCAGAGACCCCAATACC 852  
QY 845 AGAGTCAGTAAATTCAGAACTATAAACAAGCACTACAGCAGCTGGAGAAATATAAACCCCTA 904  
DB 853 AGAGTCAGTAAATTCAGAACTATAAACAAGCACTACAGCAGCTGGAGAAATATAAACCCCTA 912  
QY 905 TGAAGAAGCACTGTTTTCAGGCGAGAGCAACAGGCTGGCAGATATCAAGCATATATACGA 964  
DB 913 TGAAGAAGCACTGTTTTCAGGCGAGAGCAACAGGCTGGCAGATATCAAGCATATATACGA 972  
QY 965 TTTTGGATGAAAATTTGGCGATCTCTCGCATTCAGTTGATCTTTGAGCGGCCCTGGT 1024  
DB 973 TTTTGGATGAAAATTTGGCGATCTCTCGCATTCAGTTGATCTTTGAGCGGCCCTGGT 1032  
QY 1025 CGAGAACTGCTTGTCCAGACTTATGATCCGTTACAGTACAGTACCTAGATGCACAAC 1084  
DB 1033 CGAGAACTGCTTGTCCAGACTTATGATCCGTTACAGTACAGTACCTAGATGCACAAC 1092  
QY 1085 GAAAGTAAAGGATTTGGTATCTGTATACATACCCGCGCTATTAGAAACTGCCCTCGGAC 1144  
DB 1093 GAAAGTAAAGGATTTGGTATCTGTATACATACCCGCGCTATTAGAAACTGCCCTCGGAC 1152  
QY 1145 AGTTGCTTATGAGTCGCTACCTCTTGGCCATGGAGAGACATGGAGTTGATCATCAAGT 1204  
DB 1153 AGTTGCTTATGAGTCGCTACCTCTTGGCCATGGAGAGACATGGAGTTGATCATCAAGT 1212  
QY 1205 AATTTCTGTAACCTTCGAGAAAGCTTTGAATGCGGCTTATCCAGCCCACTGATTATGT 1264  
DB 1213 AATTTCTGTAACCTTCGAGAAAGCTTTGAATGCGGCTTATCCAGCCCACTGATTATGT 1272  
QY 1265 GGAGATTTGGCAGGCATACCTTGTATCTGAGGAGAGGTTGATTTCAACAAAGACTC 1324  
DB 1273 GGAGATTTGGCAGGCATACCTTGTATCTGAGGAGAGGTTGATTTCAACAAAGACTC 1332  
QY 1325 CAGTAAAGAGCTGGAGGAGTTGAGGCGCCCTTTACTCGCTGCTTGGAGTATCTGAAGCA 1384  
DB 1333 CAGTAAAGAGCTGGAGGAGTTGAGGCGCCCTTTACTCGCTTGGAGTATCTGAAGCA 1392  
QY 1385 GGAGTGGAGAGCGCTTTCAATCAGAGTGGTGTATCCAAAGCTCGCTGATTATGCAGAACTG 1444  
DB 1393 GGAGTGGAGAGCGCTTTCAATCAGAGTGGTGTATCCAAAGCTCGCTGATTATGCAGAACTG 1452  
QY 1445 GGCTAGGATGAGGCTCGACTGTGCAATTAACATGCAAAAGCTCGGGAACTCTGGGATAG 1504

1453	Db	 GGCTAGGATTGAGGCTCGACTGTGCAATAACATGCAGAAAGCTCGGGAACCTTGGGATAG	1511
1505	QY	 CATCATGACCAGAGGAAATGCCAAGTACGCCCAACATGTGGCTAGAGTATTACAACCTGGA	1564
1513	Db	 CATCATGACCAGAGGAAATGCCAAGTACGCCCAACATGTGGCTAGAGTATTACAACCTGGA	1572
1565	QY	 AAGAGCTCATGTGTGACACCCAGACTGCCGGAAGGCTTGCACCGGGCGTCCAGTGCAC	1624
1573	Db	 AAGAGCTCATGTGTGACACCCAGACTGCCGGAAGGCTTGCACCGGGCGTCCAGTGCAC	1632
1625	QY	 CAGTGACTACCAGAGCAGCTCTGGGAAGTGTACTCACCATGGAGAGCAGCAAGAGTTTC	1684
1633	Db	 CAGTGACTACCAGAGCAGCTCTGGGAAGTGTACTCACCATGGAGAGCAGCAAGAGTTTC	1692
1685	QY	 TTTAGAAGATTGGGATATAGCTCTTCAGAAAACTGAAACCCGATTAGCTCGTGTCAATGA	1744
1693	Db	 TTTAGAAGATTGGGATATAGCTCTTCAGAAAACTGAAACCCGATTAGCTCGTGTCAATGA	1752
1745	QY	 GCAGAGATGAAGGCTGCAGAGAAGGACGAGCCCTTGTGCAGCAAGAAAGAAAGGC	1804
1753	Db	 GCAGAGATGAAGGCTGCAGAGAAGGACGAGCCCTTGTGCAGCAAGAAAGAAAGGC	1812
1805	QY	 TGAACAACCGAAAAGAGCTCGGGCTGCAGAAAGAAAGGCTTAAAAAGAAAGAAATCAG	1864
1813	Db	 TGAACAACCGAAAAGAGCTCGGGCTGCAGAAAGAAAGGCTTAAAAAGAAAGAAATCAG	1872
1865	QY	 AGGCCCAGAGAAGCGCGGAGCAGATGAGGACGATGAGAAAGAGTGGGGCGATGATCAAGA	1924
1873	Db	 AGGCCCAGAGAAGCGCGGAGCAGATGAGGACGATGAGAAAGAGTGGGGCGATGATCAAGA	1932
1925	QY	 AGAGCAGCCTTCCAAACGAGAGGCTCGAGAACAGCATCCCTGCACTGGAGAAACACA	1984
1933	Db	 AGAGCAGCCTTCCAAACGAGAGGCTCGAGAACAGCATCCCTGCACTGGAGAAACACA	1992
1985	QY	 AAATGTAGAAGTAGCAGCAGGCGCCGCTGGGAAATGTGTCGGTATAGTGTGAGAGCCCC	2044
1993	Db	 AAATGTAGAAGTAGCAGCAGGCGCCGCTGGGAAATGTGTCGGTATAGTGTGAGAGCCCC	2052
2045	QY	 TTTGAAGCAGAGGAGGAGGCGAGCCCTCCTGTAAGAGGGACATGSCCAAGGTGCTGCACGA	2104
2053	Db	 TTTGAAGCAGAGGAGGAGGCGAGCCTCCTGTAAGAGGGACATGSCCAAGGTGCTGCACGA	2112
2105	QY	 CAGCAGCAAGGACAGCATCACCGTCTTTGTCAAGAACCTGCCCTACAGCATCGAGAGCC	2164
2113	Db	 CAGCAGCAAGGACAGCATCACCGTCTTTGTCAAGAACCTGCCCTACAGCATCGAGAGCC	2172
2165	QY	 GGACAGAGACTCAGGCCACTCTTCAGGCGCTGTGGGAGGTGGTCCAGATCGGACCCAT	2224
2173	Db	 GGACAGAGACTCAGGCCACTCTTCAGGCGCTGTGGGAGGTGGTCCAGATCGGACCCAT	2232
2225	QY	 CTTCAGCAACCGTGGGGATTTCGAGGTACTGCTACGTGGAGTTTAAAGACAGAGAAATC	2284
2233	Db	 CTTCAGCAACCGTGGGGATTTCGAGGTACTGCTACGTGGAGTTTAAAGACAGAGAAATC	2292
2285	QY	 AGCCCTTCAGGCACCTGGAGATGGACCGAAAAGTGTAGAAGGGAGGCCAATGTTGTTTC	2344
2293	Db	 AGCCCTTCAGGCACCTGGAGATGGACCGAAAAGTGTAGAAGGGAGGCCAATGTTGTTTC	2352
2345	QY	 CCCTCTGTGGATAAGACAGAAAACCCCGATTTTAAGGTGTTTCAGGTACAGCACTTCCCT	2404
2353	Db	 CCCTCTGTGGATAAGACAGAAAACCCCGATTTTAAGGTGTTTCAGGTACAGCACTTCCCT	2412
2405	QY	 AGAGAAACAACAGCTGTTTCATCTCAGGCTGCCCTTCTCTGTACTTAAAGAGAACTAGA	2464
2413	Db	 AGAGAAACAACAGCTGTTTCATCTCAGGCTGCCCTTCTCTGTACTTAAAGAGAACTAGA	2472
2465	QY	 AGAAATCTGTAAGGCTCATGGCACCGTGAAGGACCTCAGGCTGGTCAACACCGGCTGG	2524
2473	Db	 AGAAATCTGTAAGGCTCATGGCACCGTGAAGGACCTCAGGCTGGTCAACACCGGCTGG	2532
2525	QY	 CAAAACCAAGGGCCTGGCCCTACGTGGAGTATGAAATCCACGAGCGTCGACGCTGT	2584

Db	2533	CAAAACAAAGGGCCCTGGCCCTACGTGGAGTATGAAATCCAGAGCGCTGCAGAGCGTGT	2599
Qy	2585	GATGAAGATGGAGCGCATGACTATCAAAGAGAAACATCATCAAAGTGGCAATCAGCAACCC	2644
Db	2593	GATGAAGATGGAGCGCATGACTATCAAAGAGAAACATCATCAAAGTGGCAATCAGCAACCC	2652
Qy	2645	TCCTCAGAGAAAGTTCCAGAGAACCCAGAGACCAGGAAGGCACAGGTGGCCCCATGCT	2704
Db	2653	TCCTCAGAGAAAGTTCCAGAGAACCCAGAGACCAGGAAGGCACAGGTGGCCCCATGCT	2712
Qy	2705	TTTTCGCCGACACATACGAGCGAGGGGGAAGGAAGGACGACGTCTCTACTGTGCTCG	2764
Db	2713	TTTTCGCCGACACATACGAGCGAGGGGGAAGGAAGGACGACGTCTCTACTGTGCTCG	2772
Qy	2765	TGCCCTCAGCGCCCAAGTGTCTGAGCTCCTCAGGCTGAGAAGGGCCCTGCCCGGCTCC	2824
Db	2773	TGCCCTCAGCGCCCAAGTGTCTGAGCTCCTCAGGCTGAGAAGGGCCCTGCCCGGCTCC	2832
Qy	2825	TGCAGTTGGCGGCCCGACGACGCCAGCGAGGCCACCCAGAGATGCAATGCCGATTTTGC	2884
Db	2833	TGCAGTTGGCGGCCCGACGACGCCAGCGAGGCCACCCAGAGATGCAATGCCGATTTTGC	2892
Qy	2885	GCTGTTTCTGAGAAAGTGAACGGGACGCTGGGAGACAGAGAAATGCCTTACTTCACTCTG	2944
Db	2893	GCTGTTTCTGAGAAAGTGAACGGGACGCTGGGAGACAGAGAAATGCCTTACTTCACTCTG	2952
Qy	2945	CCGGCGGACCTCCACCACCCAGCAGCTGCACCTGGGATGGACAGGCGCTGGTGTGCTGG	3004
Db	2953	CCGGCGGACCTCCACCACCCAGCAGCTGCACCTGGGATGGACAGGCGCTGGTGTGCTGG	3012
Qy	3005	TGCTCGCAACCACAGATGGCTTCCTCGGCTTTTACAGAGAAAGGGGTTCTTAAGTCA	3064
Db	3013	TGCTCGCAACCACAGATGGCTTCCTCGGCTTTTACAGAGAAAGGGGTTCTTAAGTCA	3072
Qy	3065	AGACCTTTTCACTGCTCCCTCATATTGAGGGCAGTGGCAGAAAAGTGACCACCTGCGAGG	3124
Db	3073	AGACCTTTTCACTGCTCCCTCATATTGAGGGCAGTGGCAGAAAAGTGACCACCTGCGAGG	3132
Qy	3125	CTGGCCCCAGGATGTGTGTCTCCTCAGATAGTTTGTATCTTAAAGACTGAGGCACAGAAG	3184
Db	3133	CTGGCCCCAGGATGTGTGTCTCCTCAGATAGTTTGTATCTTAAAGACTGAGGCACAGAAG	3192
Qy	3185	CGAAACGAGAACACACTGTTTTTTTGAGACACAGTTGTCCAAATGTTTCTGGCCAGCTCCGG	3244
Db	3193	CGAAACGAGAACACACTGTTTTTTTGAGACACAGTTGTCCAAATGTTTCTGGCCAGCTCCGG	3252
Qy	3245	CCCCTTTTTGTATGACACTTCTCTCCACCCTGCACAGCAGTGTGCCCTCATCTCTTTT	3304
Db	3253	CCCCTTTTTGTATGACACTTCTCTCTCCACCCTGCACAGCAGTGTGCCCTCATCTCTTTT	3312
Qy	3305	AAATTTAAAGATGAAATGGCAGATGCTAGTAATTCACAGAATGGCCCTCTTGTGGGGTG	3364
Db	3313	AAATTTAAAGATGAAATGGCAGATGCTAGTAATTCACAGAATGGCCCTCTTGTGGGGTG	3372
Qy	3365	GGTCTGAGGGAAGTCAGCTATAAAACATTTTGTCTGGAGTTTGTGTTCAATGGGCTGTGCAT	3424
Db	3373	GGTCTGAGGGAAGTCAGCTATAAAACATTTTGTCTGGAGTTTGTGTTCAATGGGCTGTGCAT	3432
Qy	3425	TTTTTATATTATGCTGTTGTTAAATGACATGTGACGCCCTTGTTCATGTTTCTCTAAAGCAG	3484
Db	3433	TTTTTATATTATGCTGTTGTTAAATGACATGTGACGCCCTTGTTCATGTTTCTCTAAAGCAG	3492
Qy	3485	AATATTGCAACATTTGTTTGTATAGGAATATTCTGTGCACCTGCTGTGGACTGTTTT	3544
Db	3493	AATATTGCAACATTTGTTTGTATAGGAATATTCTGTGCACCTGCTGTGGACTGTTTT	3552
Qy	3545	CTTTGCGCTAGTACTAGTACCTGTGTGTCTTAAACATGAGTTTTCAGCCCTTTGGTTTG	3604
Db	3553	CTTTGCGCTAGTACTAGTACCTGTGTGTCTTAAACATGAGTTTTCAGCCCTTTGGTTTG	3612
Qy	3605	TTTAAATACCATGTCAAATGCAAACTCAATTCCTCCCATTTAGCTTTATTAACCTGACGT	3664
Db	3613	TTTAAATACCATGTCAAATGCAAACTCAATTCCTCCCATTTAGCTTTATTAACCTGACGT	3672

QY	3665	TCTCTTCAAACCTTCTGCTGAATGGTACTCAGATGTCAGATTCACATACACAGATGTGTTT	3724
Db	3673	TCTCTTCAAACCTTCTGCTGAATGGTACTCAGATGTCAGATTCACATACACAGATGTGTTT	3732
QY	3725	GAAGTGGGTACCTTCTTACCTTAATAGATGTGTAATAGAACTTTTCTAAGTC	3780
Db	3733	GAAGTGGGTACCTTCTTACCTTAATAGATGTGTAATAGAACTTTTCTAAGTC	3788
RESULT 2			
D63879			
LOCUS	3660 bp mRNA linear PRI 06-OCT-2001		
DEFINITION	Human mRNA for KIAA0156 gene, complete cds.		
ACCESSION	D63879.1 GI:961449		
VERSION	KIAA0156.		
KEYWORDS	Homo sapiens male "myeloblast cell_line:KG-1 cDNA to mRNA.		
SOURCE	Homo sapiens		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
REFERENCE	1 Nagase,T., Seki,N., Tanaka,A., Ishikawa,K. and Nomura,N. Prediction of the coding sequences of unidentified human genes. IV. The coding sequences of 40 new genes (KIAA0121-KIAA0160) deduced by analysis of cDNA clones from human cell line KG-1 DNA Res. 2 (4), 167-174 (1995)		
JOURNAL	612127530		
MEDLINE	2 (bases 1 to 3660)		
REFERENCE	Ohara,O., Nagase,T., Kikuno,R. and Nomura,N. Direct Submission Submitted (11-AUG-1995) Osamu Ohara, Kazusa DNA Research Institute; 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:cdnainfo@kazusa.or.jp, Tel:++81-438-52-3913)		
JOURNAL	Location/Qualifiers		
TITLE	1. ..3660		
FEATURES	/organism="Homo sapiens"		
source	/db_xref="taxon:9606"		
	/sex="male"		
	/cell_lines="KG-1"		
	/cell_type="myeloblast"		
gene	1. ..3660		
	/gene="KIAA0156"		
CDS	1. ..2892		
	/gene="KIAA0156"		
	/note="KIAA0156 gene product is related to xenopus nucleolin."		
	/codon_start=1		
	/protein_id="BAA09929.1"		
	/db_xref="GI:961450"		
	/translation="MATAAETSSAPEAESKAGPKADGEEDVEYKAAATRRKVLRSRAVA AATYTMGPAMDQEGVSGSDGEYAMASSSPGEYEWYDEEBENKQLEIERLE EQLSNVDYCHVDLIRLLEGELTKVRMARQKSEIEPLTELWLEWLHDEISMA ODGLDREHVLDLPEKAVKDXICPNWLEYGQYSGVGGIGOKGLEKVSFEERALSVG LHMTKGLWALDYREFESAIVEARLEKVLHSLFRQLAIFLYDMEATFPAYEWSDDP IPESVIQYNKALQOLEKYEKLEALLOAEAPRLAEYQAYIDEMKIGDPARLQILFE RALVENCIVLTETKALNAGFIQATDYVEIQWALDYLRVRDFKQSDSKEELELRAAFH GVDHVIQSVTEPERENSGDPCVTMWNWARIELRCNNMOKARELWDSIMTRGNAYV RALEYLRKEHRENGSCVLMQWARIELRCNNMOKARELWDSIMTRGNAYV RALWLEYLNLEHAGDHOCHKALHRAVQCTSDYPEHVCEVLLTMETEGSLEDMDIAY OKTETRLRVNEORWKAEREKAAALVQOEBEKAQRKARAEKKALKKKKIRGPKRG ADEDEKMGDDEEQFSKRRVENSIPAAGETONVEVAPAGKACAAEVEVPPSKQK EKAASLRKWDPSKSDKSIIVFVSNLPISMOEPDKLRPLFEACGEVQVIRPLFS NRKDFRGYCYVEFEKESALQALBMRKSVGEPRMFVSPCDKSKPDKFVFRYFSL EKHLFIISGLPFCSTKEBELICAKHGTVDKDLRLVTRNAGKPKGLAYVVENESQASQ AVMKMDGMTIKENIIKVAISNPPORKYPEKTPKAPGPMPLPQTYGARGKGTQLS LLPRALORPSAAPQAQENGPAAAPAAVAPAAEAPKMSNADFALFR"		
BASE COUNT	986 a	824 c	1049 g 801 t
ORIGIN			
Query Match	96.3%	Score 3656.8;	DB 9;
Best Local Similarity	99.9%;	Pred. No. 0;	
Length	3660;		

Query Match 96.3%; Score 3656.8; DB 9; Length 3660;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 3658; Conservative 0; Mismatches 2; Indels 0; Gaps 0;



1092 AAGGATTGGTTTATCTGTACATAACCGCGCTATTAGAACTGCCCTGGACAGTTGCC 1151  
1081 AAGGATTGGTTTATCTGTACATAACCGCGCTATTAGAACTGCCCTGGACAGTTGCC 1140  
1152 TTATGAGTCGGTACCTCTGTGGCCATGGAGACATGGAGTTGATCATCAAGTAATTTCT 1211  
1141 TTATGAGTCGGTACCTCTGTGGCCATGGAGACATGGAGTTGATCATCAAGTAATTTCT 1200  
1212 GTAACTTCAGAAAGCTTTGAATGCCGGCTTCCATCCAGGCCACTGATTATGTGGAGATT 1271  
1201 GTAACTTCAGAAAGCTTTGAATGCCGGCTTCCATCCAGGCCACTGATTATGTGGAGATT 1260  
1272 TGGCAGCATACCTGATTACCTGAGGAGAGAGGTTGATTTCAAACAAGACTCCAGTAA 1331  
1261 TGGCAGCATACCTGATTACCTGAGGAGAGGTTGATTTCAAACAAGACTCCAGTAA 1320  
1332 GAGCTGGAGAGTTGAGGCGCGCTTTACTCGTGCCTTGAGTATCTGAAGCAGGAGTG 1391  
1321 GAGCTGGAGAGTTGAGGCGCGCTTTACTCGTGCCTTGAGTATCTGAAGCAGGAGTG 1380  
1392 GAAGAGCGTTTCAATGAGAGTGTGATCCAAAGCTGCGTGATTATGAGAACTGGGTAGG 1451  
1381 GAAGAGCGTTTCAATGAGAGTGTGATCCAAAGCTGCGTGATTATGAGAACTGGGTAGG 1440  
1452 ATTGAGGCTGAGCTGTGCAATAACATGCAGAAAGCTCGGAACTCTGGGATAGCATATG 1511  
1441 ATTGAGGCTGAGCTGTGCAATAACATGCAGAAAGCTCGGAACTCTGGGATAGCATATG 1500  
1512 ACCAGAGGAATGCCAAGTACGCCAACATGTGCTGAGTATTACAACTGGAAAGAGCT 1571  
1501 ACCAGAGGAATGCCAAGTACGCCAACATGTGCTGAGTATTACAACTGGAAAGAGCT 1560  
1572 CATGTTGACACCCAGCACTGCCGGAAGGCTCTCACCGGCGCTGCAGTGCACCAAGTGAC 1631  
1561 CATGTTGACACCCAGCACTGCCGGAAGGCTCTGCAGCGGCGCTGCAGTGCACCAAGTGAC 1620  
1632 TACCCAGAGCAGCTCTGCGAAGTGTACTCACCATGGAGAGGACAGAAAGTTCTTTAGAA 1691  
1621 TACCCAGAGCAGCTCTGCGAAGTGTACTCACCATGGAGAGGACAGAAAGTTCTTTAGAA 1680  
1692 GATTGGATAGCTGTTCAAGAACTGAAACCCGATTAGCTGTTCAATGAGCAGAGA 1751  
1681 GATTGGATAGCTGTTCAAGAACTGAAACCCGATTAGCTGTTCAATGAGCAGAGA 1740  
1752 ATGAAGGCTCAGAGAGGAGCAGCCCTTGTGCAGCAAGAAAGAAAGGCTGAAACAA 1811  
1741 ATGAAGGCTCAGAGAGGAGCAGCCCTTGTGCAGCAAGAAAGAAAGGCTGAAACAA 1800  
1812 CGAAAGAGCTCGGGCTGAGAGAAAGCGTTTAAAAAGAAAGAAAGATCAGAGGCCCA 1871  
1801 CGAAAGAGCTCGGGCTGAGAGAAAGCGTTTAAAAAGAAAGAAAGATCAGAGGCCCA 1860  
1872 GAGAGCGGAGCAGATGAGGACGATGAGAAAGATGGGGCGATGATGAAGAGAGCAG 1931  
1861 GAGAGCGGAGCAGATGAGGACGATGAGAAAGATGGGGCGATGATGAAGAGAGCAG 1920  
1932 CCTTCCAAACGCAAGAGGTCGAGAACAGCATCCCTGCAGCTGGAGAAACACAAATGTA 1991  
1921 CCTTCCAAACGCAAGAGGTCGAGAACAGCATCCCTGCAGCTGGAGAAACACAAATGTA 1980  
1992 GAAGTAGCAGAGGCGCGCTGGGAAATGTGCTGCGGTAGATGTGAGGCCCTTCGAAG 2051  
1981 GAAGTAGCAGAGGCGCGCTGGGAAATGTGCTGCGGTAGATGTGAGGCCCTTCGAAG 2040  
2052 CAGAGGAGAGGAGCAGCTCCCTGAGAGGAGCATGCCAAGGTGCTGCAGCAGCAGC 2111  
2041 CAGAGGAGAGGAGCAGCTCCCTGAGAGGAGCATGCCAAGGTGCTGCAGCAGCAGC 2100  
2112 AAGGACAGCATCCGCTTTGTGACCAACTCCCTTACAGCATGAGGAGCGGACAGC 2171  
2101 AAGGACAGCATCCGCTTTGTGACCAACTCCCTTACAGCATGAGGAGCGGACAGC 2160  
2172 AAGCTAGGCGCACTCTTCGAGGCGCTGTGGGAGGTTGGTCCAGATCCGACCCATCTTCAGC 2231

2161 AAGCTCAGGCGCACTCTTCGAGGCGCTGTGGGAGGTTGGTCCAGATCCGACCATCTTCAGC 2220  
2232 AACCGTGGGATTTCCGAGGTTACTGCTAGCTGAGGATTTAAAGAGAGAAATCAGCCCTT 2291  
2221 AACCGTGGGATTTCCGAGGTTACTGCTAGCTGAGGATTTAAAGAGAGAAATCAGCCCTT 2280  
2292 CAGGCACTGAGATGGAACGAAAGTGTAGAGGAGGCAATGTTTGTTCCTCCCTGT 2351  
2281 CAGGCACTGAGATGGAACGAAAGTGTAGAGGAGGCAATGTTTGTTCCTCCCTGT 2340  
2352 GTGGATAAGAGCAAAAAACCCGATTTTAAAGTGTTCAGGTACAGCACTTCCCTAGAGAAA 2411  
2341 GTGGATAAGAGCAAAAAACCCGATTTTAAAGTGTTCAGGTACAGCACTTCCCTAGAGAAA 2400  
2412 CACAAGCTGTTCATCTCAGGCGCTGCCCTTCTCCTGTACTAAAGAGGAACTAGAGAAATC 2471  
2401 CACAAGCTGTTCATCTCAGGCGCTGCCCTTCTCCTGTACTAAAGAGGAACTAGAGAAATC 2460  
2472 TGTAAAGGCTCATGCGACCGTGAAGGACCTCAGGCTGTGTCAACACCGGCTGGCAACCA 2531  
2461 TGTAAAGGCTCATGCGACCGTGAAGGACCTCAGGCTGTGTCAACACCGGCTGGCAACCA 2520  
2532 AAGGCGCTGCGCTACGTGGAGTATGAAATGAATCCAGGCGTGCAGGCTGTGATGAAG 2591  
2521 AAGGCGCTGCGCTACGTGGAGTATGAAATGAATCCAGGCGTGCAGGCTGTGATGAAG 2580  
2592 ATGAGCGGATGATCAATCAAGAGAACATCATCAAGTGGCAATCAGCAACCCCTCCAG 2651  
2581 ATGAGCGGATGATCAATCAAGAGAACATCATCAAGTGGCAATCAGCAACCCCTCCAG 2640  
2652 AGAAAGTTCAGAGAAAGCAGAGACCCAGGAGCAGGTCAGGCTGGCGGCTGCTTTGCCG 2711  
2641 AGAAAGTTCAGAGAAAGCAGAGACCCAGGAGCAGGTCAGGCTGGCGGCTGCTTTGCCG 2700  
2712 CAGACATACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2771  
2701 CAGACATACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2760  
2772 CAGCGCCCAAGTCTGCGCTCCCTCAGGCTGAGAACCGGCTGCCGCGGCTCCCTGCGAGTT 2831  
2761 CAGCGCCCAAGTCTGCGCTCCCTCAGGCTGAGAACCGGCTGCCGCGGCTCCCTGCGAGTT 2820  
2832 GCGCGCCCAAGTCTGCGCTCCCTCAGGCTGAGAACCGGCTGCCGCGGCTCCCTGCGAGTT 2891  
2821 GCGCGCCCAAGTCTGCGCTCCCTCAGGCTGAGAACCGGCTGCCGCGGCTCCCTGCGAGTT 2880  
2892 CTGAGAAAGTGAACGGGAGCCTGGGAGACAGGAAATGCCCTTACTTACTCTGCGCCGCG 2951  
2881 CTGAGAAAGTGAACGGGAGCCTGGGAGACAGGAAATGCCCTTACTTACTCTGCGCCGCG 2940  
2952 GAGCTCCCAAGCAGCAGCAGTGCACCTGGGATGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3011  
2941 GAGCTCCCAAGCAGCAGCAGTGCACCTGGGATGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3000  
3012 AACACAGATGGCTCTCGGCTTTAGACAGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3071  
3001 AACACAGATGGCTCTCGGCTTTAGACAGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3060  
3072 TTGAGTGTCTCTCATATTTAGGAGCAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3131  
3061 TTGAGTGTCTCTCATATTTAGGAGCAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3120  
3132 CAGGATGTGTCTCTCATATTTAGGAGCAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3191  
3121 CAGGATGTGTCTCTCATATTTAGGAGCAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3180  
3192 AGAACACACTGTTTGTGAGACAGTGTGCTCAATGTTTCTGCGGAGCTCGGCGCCCTTT 3251  
3181 AGAACACACTGTTTGTGAGACAGTGTGCTCAATGTTTCTGCGGAGCTCGGCGCCCTTT 3240  
3252 TTGATGACACTTCTCTTCCACCTTGACAGCAGATGTGCCGCTCATTTTAAATTTTA 3311



968	TTTTGAGATGAATAATTGGCGATCCTGCTGCATTCAGTTGATCTTTGAGCGCCCTGGT	1027	2105	CAGCAACAAGGACAGCATCACCGTCTTTTGTTCAGCAACCTGCCCTACAGCATGCGAGAGCC	2164
1025	CGAGAACTGCTTGTCCAGACTTATGGATCCGTTACAGTCAGTCAGTACCTAGATCGACAACT	1084	2097	CAGCAGCAAGGACAGCATCACCGTCTTTGTTCAGCAACCTGCCCTACAGCATGCGAGAGCC	2156
1028	CGAGAACTGCTTGTCCAGACTTATGGATCCGTTAC-----AGATCGACAACCT	1076	2165	GGACAGAACTCAGGCCACTCTTCGAGGCGCTGTGGGAGGTGGTTCAGATCCGACCCCAT	2224
1085	GAAGTAAAGGATTTGGTTTTTATCTGTGTACATAACCGCGCTATTAGAACTGCCCTGGAC	1144	2157	GGACAGAACTCAGGCCACTCTTCGAGGCGCTGTGGGAGGTGGTTCAGATCCGACCCCAT	2216
1077	GAAGTAAAGGATTTGGTTTTTATCTGTGTACATAACCGCGCTATTAGAACTGCCCTGGAC	1136	2225	CTTCACAACCGTGGGATTTCCGAGGTTACTGTACGTGGAGTTTAAAGAGAGAAATC	2284
1145	AGTTGCCCTATGAGTCGGTACCTCTTGGCCATGGAGAGACATGGAGTTGATCATCAAGT	1204	2217	CTTCACAACCGTGGGATTTCCGAGGTTACTGTACGTGGAGTTTAAAGAGAGAAATC	2276
1137	AGTTGCCCTATGAGTCGGTACCTCTTGGCCATGGAGAGACATGGAGTTGATCATCAAGT	1196	2285	AGCCCTTCAGGCACTGGAGATGGACCGGAAAGCTGTAGAAGGGAGGCGCAATGTTTGTTC	2344
1205	AAATTCGTAACTTCGAGAAAGCTTTGAATCGCGCTTCATCCAGGCCACTGATTATGT	1264	2277	AGCCCTTCAGGCACTGGAGATGGACCGGAAAGCTGTAGAAGGGAGGCGCAATGTTTGTTC	2336
1197	AAATTCGTAACTTCGAGAAAGCTTTGAATCGCGCTTCATCCAGGCCACTGATTATGT	1256	2345	CCCTGTGTGATTAAGAGCAAAAACCCGATTTTAAAGTGTTCAGGTACAGCACTTCCCT	2404
1265	GGAGATTTGGCAGGCATPACCTTGATTACCTTGAGGAGAAAGGTTGATTCAAAACAAGCTC	1324	2337	CCCTGTGTGATTAAGAGCAAAAACCCGATTTTAAAGTGTTCAGGTACAGCACTTCCCT	2396
1257	GGAGATTTGGCAGGCATPACCTTGATTACCTTGAGGAGAAAGGTTGATTCAAAACAAGCTC	1316	2405	AGAGAAACACAGCTGTTTATCTCAGGCCCTGCCCTTTCTCCTGTACTTAAAGAGAACTAGA	2464
1325	CAGTAAAGACCTGGAGAGTTGAGGCGCGCTTTACTCGTGCCTGGAGTATCTGAAGCA	1384	2397	AGAGAAACACAGCTGTTTATCTCAGGCCCTGCCCTTTCTCCTGTACTTAAAGAGAACTAGA	2456
1317	CAGTAAAGACCTGGAGAGTTGAGGCGCGCTTTACTCGTGCCTGGAGTATCTGAAGCA	1376	2465	AGAAATCTGTAAGGCTCATGGCACCGTGAAGGACCTCAGGCTGGTCACCAACCGGCTGG	2524
1385	GGAGTGGAAAGAGCGTTTCAATGAGAGTGTGTATCAAGCTGCGTATTATGAGAACTG	1444	2457	AGAAATCTGTAAGGCTCATGGCACCGTGAAGGACCTCAGGCTGGTCACCAACCGGCTGG	2516
1377	GGAGTGGAAAGAGCGTTTCAATGAGAGTGTGTATCAAGCTGCGTATTATGAGAACTG	1436	2525	CAAAACAAAGGGCTTGGCCTTACGTGGAGTATGAAATGAATCCAGGCGCTCGCAGGCTGT	2584
1445	GGCTAGGATTTAGGCTCGACTGTGCAATAACATGCAGAAAGCTCGGGAACCTCTGGGTAG	1504	2517	CAAAACAAAGGGCTTGGCCTTACGTGGAGTATGAAATGAATCCAGGCGCTCGCAGGCTGT	2576
1437	GGCTAGGATTTAGGCTCGACTGTGCAATAACATGCAGAAAGCTCGGGAACCTCTGGGTAG	1496	2585	GATGAAGATGGACGGCATGACTATCAAGAGAACATCATCAAAAGTGGCAATCAGCAACCC	2644
1505	CATCATGACAGAGGAAATGCCAAGTAGGCCACACATGTGGCTAGATTTACAACTGGA	1564	2577	GATGAAGATGGACGGCATGACTATCAAGAGAACATCATCAAAAGTGGCAATCAGCAACCC	2636
1497	CATCATGACAGAGGAAATGCCAAGTAGGCCACACATGTGGCTAGATTTACAACTGGA	1556	2645	TCCTCAGAGGAAAGTTCCAGAGAAAGCCAGAGACCCAGAGGACACAGGTGGCCCATGCT	2704
1565	AAGAGCTCATGTTGACACCCAGACTGCGCGAAGGCTCTCACCGGCGCTCCAGTGCAC	1624	2637	TCCTCAGAGGAAAGTTCCAGAGAAAGCCAGAGACCCAGAGGACACAGGTGGCCCATGCT	2696
1557	AAGAGCTCATGTTGACACCCAGACTGCGCGAAGGCTCTCACCGGCGCTCCAGTGCAC	1616	2705	TTTTGCCGCAGACATACGGAGCGAGGGGAAAGGAGGACGACGTCTCTACTGCTCG	2764
1625	CAGTGACTACCCAGAGCAGCTCTGCGAAGTGTACTCACCATGGAGGAGACAGAGTTTC	1684	2697	TTTTGCCGCAGACATACGGAGCGAGGGGAAAGGAGGACGACGTCTCTACTGCTCG	2756
1617	CAGTGACTACCCAGAGCAGCTCTGCGAAGTGTACTCACCATGGAGGAGACAGAGTTTC	1676	2765	TGCCCTGACAGCGCCCAAGTGTCTCAGGCTGAGAACGGCCCTGCGCGGCGTCC	2824
1685	TTTAGAAGATTTGGATATAGCTGTTCAAGAACTGAAACCCGATTAAGTGTCAATGA	1744	2757	TGCCCTGACAGCGCCCAAGTGTCTCAGGCTGAGAACGGCCCTGCGCGGCGTCC	2816
1677	TTTAGAAGATTTGGATATAGCTGTTCAAGAACTGAAACCCGATTAAGTGTCAATGA	1736	2825	TGCAGTTGCCGCCCCAGCAGCCACCGAGGACCCCAAGATGTCCAATGCCGATTTGCCAA	2884
1745	GCAGAGATGAAGCTGCAGAGAGGAGGAGCAGCCCTGTGCAGCAAGAGAAAGGCG	1804	2817	TGCAGTTGCCGCCCCAGCAGCCACCGAGGACCCCAAGATGTCCAATGCCGATTTGCCAA	2876
1737	GCAGAGATGAAGCTGCAGAGAGGAGGAGCAGCCCTGTGCAGCAAGAGAAAGGCG	1796	2885	GCTGTTTCTGAGAAAGTGAACCGGACGCTGGGAGACAGGAAATGSCCTTACTTCACCTCTGG	2944
1805	TGAACAACGAAAGAGCTCGGGCTGAGAGAAAGCGTTAAAAAGAGAAAGATCAG	1864	2877	GCTGTTTCTGAGAAAGTGAACCGGACGCTGGGAGACAGGAAATGSCCTTACTTCACCTCTGG	2936
1797	TGAACAACGAAAGAGCTCGGGCTGAGAGAAAGCGTTAAAAAGAGAAAGATCAG	1856	2945	CCGCGGGAGCTCCACACCCAGCAGTGCATGGGGATGGACAGGCTTGGTGTCTGGG	3004
1865	AGGCCAGAGAGCGGACAGATGAGGACGATGAGAAAGTGGGGCGATGATGAAGA	1924	2937	CCGCGGGAGCTCCACACCCAGCAGTGCATGGGGATGGACAGGCTTGGTGTCTGGG	2996
1857	AGGCCAGAGAGCGGACAGATGAGGACGATGAGAAAGTGGGGCGATGATGAAGA	1916	3005	TGCTCCCAACACAGATGGCTCCTCGGCTTTAGACAGAAAGGGGAGGTTCTTAAGTCA	3064
1925	AGAGACGCTTCCAAACGAGAGGTCGAGACAGCATCCCTGCAGCTGGAGAAACACA	1984	2997	TGCTCCCAACACAGATGGCTCCTCGGCTTTAGACAGAAAGGGGAGGTTCTTAAGTCA	3056
1917	AGAGACGCTTCCAAACGAGAGGTCGAGACAGCATCCCTGCAGCTGGAGAAACACA	1976	3065	AGAGCCTTTTCACTGCTCCCTCATATTGAGGCGAGTGGCAGAAAGTGACCACTCTCGAGG	3124
1985	AAATGTAGAAGTAGCAGCAGGCGCGCTGGGAAATGTGCTCCGCTAGATGTGGAGCCCC	2044	3057	AGAGCCTTTTCACTGCTCCCTCATATTGAGGCGAGTGGCAGAAAGTGACCACTCTCGAGG	3116
1977	AAATGTAGAAGTAGCAGCAGGCGCGCTGGGAAATGTGCTCCGCTAGATGTGGAGCCCC	2036	3125	CTGGGCCCCAGGATGTGGTGTCTGATAGTTTTTGTATCTTAAAGACTGAGGCACAGAAG	3184
2045	TTGAAACAGAGAGGAGCAGCTCCCTGAAAGAGGACATGCCCAAGGTGTGCACGA	2104	3117	CTGGGCCCCAGGATGTGGTGTCTGATAGTTTTTGTATCTTAAAGACTGAGGCACAGAAG	3176
2037	TTGAAACAGAGAGGAGCAGCTCCCTGAAAGAGGACATGCCCAAGGTGTGCACGA	2096			

```
QY 3185 CGAAACGAGAACACACTGTTTTGTGACACAGTGTGCCAAATGTTTCTGGCCAGCTCCGG 3244
Db 3177 CGAAACGAGAACACACTGTTTTGTGACACAGTGTGCCAAATGTTTCTGGCCAGCTCCGG 3236
QY 3245 CCCCTTTTGTATGACACACTTCTTCCACCCTGCACAGCACATGTGCCCGTCAATCTTTT 3304
Db 3237 CCCCTTTTGTATGACACACTTCTTCCACCCTGCACAGCACATGTGCCCGTCAATCTTTT 3296
QY 3305 AATTTTAAAGATGAATGGCAGATCTAGTAAATTCACAGAATGGCTCTGTGGGGTG 3364
Db 3297 AATTTTAAAGATGAATGGCAGATCTAGTAAATTCACAGAATGGCTCTGTGGGGTG 3356
QY 3365 GGTCTGAGGGAAGTCAGCTATAAAACATTTGCTGGAGTTTGTTCATGSGGCTGTGCAT 3424
Db 3357 GGTCTGAGGGAAGTCAGCTATAAAACATTTGCTGGAGTTTGTTCATGSGGCTGTGCAT 3416
QY 3425 TTTTATATATGTTTGTAAATGACATCTACGCCCTTGTTCATGTTTCTTAAAGCAG 3484
Db 3417 TTTTATATATGTTTGTAAATGACATCTACGCCCTTGTTCATGTTTCTTAAAGCAG 3476
QY 3485 AATATTGCAACATTTGTTTGTATAGGAATTTTGTGCCACCTGCTGTGACTGTTTT 3544
Db 3477 AATATTGCAACATTTGTTTGTATAGGAATTTTGTGCCACCTGCTGTGACTGTTTT 3536
QY 3545 CTTTGCTAGTACAGTACCTGTTGCTCTAAACATGAGTTTCAGCCCTTTGGTTTG 3604
Db 3537 CTTTGCTAGTACAGTACCTGTTGCTCTAAACATGAGTTTCAGCCCTTTGGTTTG 3596
QY 3605 TTTAATPACCATGTCAATCAAACTCAATCTCCCATTTAGCTTTA 3652
Db 3597 TTTAATPACCATGTCAATCAAACTCAATCTCCCATTTAGCTTTA 3644

RESULT 4
AX268309
LOCUS AX268309 3537 bp DNA linear PAT 29-OCT-2001
DEFINITION Sequence 30 from Patent WO0174855.
ACCESSION AX268309
VERSION AX268309.1 GI:16541547
KEYWORDS synthetic construct.
SOURCE synthetic construct.
ORGANISM artificial sequences.
REFERENCE
1
Laus, R., Vidovic, D. and Graddis, T.
Compositions and methods for dendritic cell-based immunotherapy
Patent: WO 0174855-A 30 11-OCT-2001;
DENDREON CORPORATION (US)
FEATURES
Location/Qualifiers
source
1..3537
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="SART-3-1C"
BASE COUNT 931 a 885 c 1048 g 673 t
ORIGIN
Query Match 76.0%; Score 2886; DB 6; Length 3537;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2886; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 12 ATGGGACTGCGGCGGAACCTTCGGCTTCAGAACCCGAGGCTGAGTCCAGGCTGGGCC 71
Db 1 ATGGGACTGCGGCGGAACCTTCGGCTTCAGAACCCGAGGCTGAGTCCAGGCTGGGCC 60
QY 72 AAGGCTGAGGAGAGGAGTGAAGTGAAGCGCTAGGACAGGAAGGTATGCG 131
Db 61 AAGGCTGAGGAGAGGAGTGAAGTGAAGCGCTAGGACAGGAAGGTATGCG 120
QY 132 CGGGCTGTGGCGCTGCGACATACAGACCATGGGGCCAGCGCTGGGATCAGCAGAGGAA 191
Db 121 CGGGCTGTGGCGCTGCGACATACAGACCATGGGGCCAGCGCTGGGATCAGCAGAGGAA 180
QY 192 GCGTGTAGGAGAGGAGTGGGATGAGTACGCCATGGCTTCCTCCGGGAGAGCTCCCCC 251
```

```
Db 181 GGCGTGAGCGAGCGATGGGATGAGTACGCCATGGCTCTCTCCCGGAGAGCTCCCC 240
QY 252 GGGGAGTACAGTGGGAATATGACGAAGAGGAGGAGAAAACACAGCTGAGATTGAGAGA 311
Db 241 GGGGAGTACAGTGGGAATATGACGAAGAGGAGGAGAAAACACAGCTGAGATTGAGAGA 300
QY 312 CTGGAGGACAGTGTCTATCAACGTCTATGACTACAACTGCGCATGTGACTTGATCAGA 371
Db 301 CTGGAGGACAGTGTCTATCAACGTCTATGACTACAACTGCGCATGTGACTTGATCAGA 360
QY 372 CTGCTCAGCTGGAAAGGGAGCTTACCAAGGTGAGGATGCCCGCCAGAGATGAGTGAA 431
Db 361 CTGCTCAGCTGGAAAGGGAGCTTACCAAGGTGAGGATGCCCGCCAGAGATGAGTGAA 420
QY 432 ATCTTTCCCTTCACTGAGAGCTCTGGCTGGAGTGGCTGCATGACAGATCAGCATGGCC 491
Db 421 ATCTTTCCCTTCACTGAGAGCTCTGGCTGGAGTGGCTGCATGACAGATCAGCATGGCC 480
QY 492 CAGGATGGCTGGACAGAGAGCAGCTGTATGACCTCTTTGAGAAAACCCGTGAAGATTAC 551
Db 481 CAGGATGGCTGGACAGAGAGCAGCTGTATGACCTCTTTGAGAAAACCCGTGAAGATTAC 540
QY 552 ATTGTCTTAACATTTGGCTAGAGTATGGCCAGTACTCAAGTGGTGGGATTTGGTCAGAA 611
Db 541 ATTGTCTTAACATTTGGCTAGAGTATGGCCAGTACTCAAGTGGTGGGATTTGGTCAGAA 600
QY 612 GGTGGCTTTGAGAAAGTTGCTCGCTGTTTGAAGGGCTCTCTCGTCTCTGTTTACAT 671
Db 601 GGTGGCTTTGAGAAAGTTGCTCGCTGTTTGAAGGGCTCTCTCGTCTCTGTTTACAT 660
QY 672 ATGACCAAAGACTCGCCCTCTGGAGGCTTACCAGAGATTTGAAAGTCGATTTGTGAA 731
Db 661 ATGACCAAAGACTCGCCCTCTGGAGGCTTACCAGAGATTTGAAAGTCGATTTGTGAA 720
QY 732 GCTGCTGGCTTGAGAAAGTCCACAGTCTTTTCCGGCGACAGTGGCGATCCCACTCAT 791
Db 721 GCTGCTGGCTTGAGAAAGTCCACAGTCTTTTCCGGCGACAGTGGCGATCCCACTCAT 780
QY 792 GATATGGAGGCCACATTTGCAGAGTATGAAGATGGTCAAGAACCAATACACAGATCA 851
Db 781 GATATGGAGGCCACATTTGCAGAGTATGAAGATGGTCAAGAACCAATACACAGATCA 840
QY 852 GTAATTCAGAACTATAACAAAGCACTACAGAGCTGGAGAAATATAAAACCCCTATGAAG 911
Db 841 GTAATTCAGAACTATAACAAAGCACTACAGAGCTGGAGAAATATAAAACCCCTATGAAG 900
QY 912 GCACCTGTTGCAGGAGAGGACCAAGGCTGGCAGAAATATCAAGCATATATCATTTTGAG 971
Db 901 GCACCTGTTGCAGGAGAGGACCAAGGCTGGCAGAAATATCAAGCATATATCATTTTGAG 960
QY 972 ATGAAAATTTGGGATCCCTGCTCGCATTCAGTTTGTATCTTTGAGCGGCCCTGTCAGAA 1031
Db 961 ATGAAAATTTGGGATCCCTGCTCGCATTCAGTTTGTATCTTTGAGCGGCCCTGTCAGAA 1020
QY 1032 TGCCTTTGCCAGACTTTATGATCCGCTTACAGTCACTAGTACAGTCACTAGTCACTAGT 1091
Db 1021 TGCCTTTGCCAGACTTTATGATCCGCTTACAGTCACTAGTCACTAGTCACTAGTCACTAG 1080
QY 1092 AAGGATTTGGTTTTTATCTGTACATACCCGCTATTTAGAACTGCCCTGGACAGTTGCC 1151
Db 1081 AAGGATTTGGTTTTTATCTGTACATACCCGCTATTTAGAACTGCCCTGGACAGTTGCC 1140
QY 1152 TTATGAGTTCGCTACCTCTTGGCCATGGAGACATGGAGTTGATCATCAAGTAAATTTCT 1211
Db 1141 TTATGAGTTCGCTACCTCTTGGCCATGGAGACATGGAGTTGATCATCAAGTAAATTTCT 1200
QY 1212 GTAACTTTGAGAAAGCTTTGAATCCGCTTTATCCAGGCCACTGATTTATGTGGAGATT 1271
Db 1201 GTAACTTTGAGAAAGCTTTGAATCCGCTTTATCCAGGCCACTGATTTATGTGGAGATT 1260
QY 1272 TGGCAGGCTATACCTTTGATTTACCTGAGGAGAGGGTTGATTTCAACAACAGACTCCAGTAA 1331
```

Db	1261	TGGCAGGCAATACCTTGATTACCTGAGGAGAAGGGTTGATTTCAAACAGACTCCAGTAAA	1320
Qy	1332	GAGCTGGAGGAGTTGAGGGCCGCTTTTACTCGTCCCTTGAGTATCTGAAGCAGGAGGTG	1391
Db	1321	GAGCTGGAGGAGTTGAGGGCCGCTTTTACTCGTCCCTTGAGTATCTGAAGCAGGAGGTG	1380
Qy	1392	GAAGAGCGTTTCAATAGAGAGTGGTGATCCAAAGCTGCGTGATTATGCAGAACTGGGGTAGG	1451
Db	1381	GAAGAGCGTTTCAATAGAGAGTGGTGATCCAAAGCTGCGTGATTATGCAGAACTGGGGTAGG	1440
Qy	1452	ATTGAGCGTCACTGTGCAATACATCCAGAAAGCTCGGAACTCTGGGATAGCATATG	1511
Db	1441	ATTGAGCGTCACTGTGCAATACATCCAGAAAGCTCGGAACTCTGGGATAGCATATG	1500
Qy	1512	ACCAGAGAAATGCCAAGTACGCCCAACATGTGGCTAGAGTATTACAACCTTGGAAAGAGCT	1571
Db	1501	ACCAGAGAAATGCCAAGTACGCCCAACATGTGGCTAGAGTATTACAACCTTGGAAAGAGCT	1560
Qy	1572	CATGGTGACACCCAGCACTGCCGGAAGGCTCTGCACCCGGGCGCTCCAGTGCACCAAGTGAC	1631
Db	1561	CATGGTGACACCCAGCACTGCCGGAAGGCTCTGCACCCGGGCGCTCCAGTGCACCAAGTGAC	1620
Qy	1632	TACCACAGGACGCTCTCGGAGTGTACTCACCATGAGAGGACAGAGGTTCTTTAGAA	1691
Db	1621	TACCACAGGACGCTCTCGGAGTGTACTCACCATGAGAGGACAGAGGTTCTTTAGAA	1680
Qy	1692	GATTGGGATATAGCTGTTCAAGAAACTGAAACCCGATTAGCTCGTGTCAATGAGCAGAGA	1751
Db	1681	GATTGGGATATAGCTGTTCAAGAAACTGAAACCCGATTAGCTCGTGTCAATGAGCAGAGA	1740
Qy	1752	ATGAAGCGTGCAGAGAAGGAGCCCTTGTGCAGCAAGAAAGAAAGAGGCTTGAACAA	1811
Db	1741	ATGAAGCGTGCAGAGAAGGAGCCCTTGTGCAGCAAGAAAGAAAGAGGCTTGAACAA	1800
Qy	1812	CGGAAAGAGCTCGGCTGAGAGAAAGCGTTAAAAAGAGAGAAAGATCAGAGGCCCA	1871
Db	1801	CGGAAAGAGCTCGGCTGAGAGAAAGCGTTAAAAAGAGAGAAAGATCAGAGGCCCA	1860
Qy	1872	GAGAAGCGGAGCAGATGAGGAGCGATGAGAAAGAGTGGGCGGATGATGAAGAGAGCAG	1931
Db	1861	GAGAAGCGGAGCAGATGAGGAGCGATGAGAAAGAGTGGGCGGATGATGAAGAGAGCAG	1920
Qy	1932	CTTTCCAAACGACGAGGGTGGAGACAGCATCCCTGCAGTGGAGAAACACAAAATGTA	1991
Db	1921	CTTTCCAAACGACGAGGGTGGAGACAGCATCCCTGCAGTGGAGAAACACAAAATGTA	1980
Qy	1992	GAAGTACGACGAGGCGCGTGGGAAATGTCTCCGCTAGATGTGGAGCCCTTCCAAG	2051
Db	1981	GAAGTACGACGAGGCGCGTGGGAAATGTCTCCGCTAGATGTGGAGCCCTTCCAAG	2040
Qy	2052	CAGAAGGAGAGGCGAGCTCCCTGAAGAGGACATGCCAAGGTGCTGCACGACGACGAGC	2111
Db	2041	CAGAAGGAGAGGCGAGCTCCCTGAAGAGGACATGCCAAGGTGCTGCACGACGACGAGC	2100
Qy	2112	AAGCAGACATCACCGCTTTGTGACCAACCTGCGCCCTACAGCATGCGAGGCGCGGACAG	2171
Db	2101	AAGCAGACATCACCGCTTTGTGACCAACCTGCGCCCTACAGCATGCGAGGCGCGGACAG	2160
Qy	2172	AAGCTAGGCGCACTCTTCGAGGCGCTGTGGGAGGTGGTCCAGATCCGACCATCTTCAGC	2231
Db	2220	AAGCTAGGCGCACTCTTCGAGGCGCTGTGGGAGGTGGTCCAGATCCGACCATCTTCAGC	2220
Qy	2232	AACCGTGGGATTTCCGAGGTTACTGCTACGTGAGGTTTAAAGAGAGAAATCAGCCCTT	2291
Db	2221	AACCGTGGGATTTCCGAGGTTACTGCTACGTGAGGTTTAAAGAGAGAAATCAGCCCTT	2280
Qy	2292	CAGGCACTGGAGATGGACCGGAAAGTGTAGAAGGAGGCGCAATGTTGTTTCCCTCGT	2351
Db	2281	CAGGCACTGGAGATGGACCGGAAAGTGTAGAAGGAGGCGCAATGTTGTTTCCCTCGT	2340
Qy	2352	GTGGATAGACAAAACCCCGATTTTAAGGTGTTAGGTACACACCTTCCCTAGAGAAA	2411
Db	2341	GTGGATAGACAAAACCCCGATTTTAAGGTGTTAGGTACACACCTTCCCTAGAGAAA	2400
Qy	2412	CACAAGCTGTTCACTCTCAGGCTGCCCTTCTCCTGTACTAAAGAGGAAGTAAAGAAATC	2471
Db	2401	CACAAGCTGTTCACTCTCAGGCTGCCCTTCTCCTGTACTAAAGAGGAAGTAAAGAAATC	2460
Qy	2472	TGTAAGGCTCATGACACCGTGAAGGACCTCAGGCTGGTGCACCAACCGGCTGGCAACCA	2531
Db	2461	TGTAAGGCTCATGACACCGTGAAGGACCTCAGGCTGGTGCACCAACCGGCTGGCAACCA	2520
Qy	2532	AAGGCGCTGCGCTACGTGGAGTATGAATAATCCAGGCGTGCAGGCTGTGATGAAG	2591
Db	2521	AAGGCGCTGCGCTACGTGGAGTATGAATAATCCAGGCGTGCAGGCTGTGATGAAG	2580
Qy	2592	ATGACCGGCTGACTATCAAGAGAAATCATCAAAAGTGGCAATCAGCAACCCCTCCCTCAG	2651
Db	2581	ATGACCGGCTGACTATCAAGAGAAATCATCAAAAGTGGCAATCAGCAACCCCTCCCTCAG	2640
Qy	2652	AGGAAAGTTCAGAGAAGCCAGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	2711
Db	2641	AGGAAAGTTCAGAGAAGCCAGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	2700
Qy	2712	CAGACATACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	2771
Db	2701	CAGACATACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	2760
Qy	2772	CAGGCGCCCAAGTGTGAGCTCCTCAGGCTGAGAAAGGCGCTGCGGCGGCTCCTGCGAGTT	2831
Db	2761	CAGGCGCCCAAGTGTGAGCTCCTCAGGCTGAGAAAGGCGCTGCGGCGGCTCCTGCGAGTT	2820
Qy	2832	GCCGCGCCAGCAGCAGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	2891
Db	2821	GCCGCGCCAGCAGCAGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	2880
Qy	2892	CTGAGA 2897	
Db	2881	CTGAGA 2886	
RESULT 5			
AF172722			
LOCUS	AF172722	3586 bp	mRNA linear ROD 24-APR-2000
DEFINITION	Mus musculus tumor-rejection antigen SART3 (Sart3) mRNA, complete cds.		
ACCESSION	AF172722		
VERSION	AF172722.1	GI:7637844	
KEYWORDS			
SOURCE	Mus musculus.		
ORGANISM	Mus musculus		
REFERENCE	1 (bases 1 to 3586)		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS	Harada, K., Yamada, A., Mine, T., Kawagoe, N., Takasu, H. and Itoh, K.		
TITLE	Mouse homologue of the human SART3 gene encoding tumor-rejection antigen		
JOURNAL	Jpn. J. Cancer Res. 91 (2), 239-247 (2000)		
MEDLINE	2022349		
PUBMED	10761712		
REFERENCE	2 (bases 1 to 3586)		
AUTHORS	Harada, K., Yamada, A., Mine, T., Kawagoe, N., Takasu, H. and Itoh, K.		
TITLE	Direct Submission		
JOURNAL	Submitted (28-JUL-1999) Immunology, Kurume University School of Medicine, 67-Asahi-Machi, Kurume, Fukuoka 830-0011, Japan		
FEATURES	Location/Qualifiers		
Source	1. 3586		
	/organism="Mus musculus"		
	/strain="BALB/c"		
	/db_xref="taxon:10090"		
	/chromosome="5"		
	/map="between D5Mit317 and D5Mit158"		
	/tissue.type="squamous cell carcinoma"		
	/dev_stage="9-10 weeks"		
	1. .3586		
	/gene="Sart3"		
gene			









Number of N's in consensus :

20

7218 ATATTGAGC

AGTGGCAGAAAGTGACCACTCTGCAGGCTGGGCCCAGGATGTGG

Morgan, M., Morris, S., Nash, S., Nelson, A., Nguyen, K., Nguyen, N.,  
Nguyen, S., Oswal, G., Parish, B., Paxton, S., Paxton, B., Perez, L.,

Scherer, S., Shah, E., Shen, H., Shlim, C., Simon, M., Sparks, A.,  
Stamps, A., Sugcang, R., Tabor, P., Taylor, T., Vasquez, L., Vinson, R.,  
Vo, Q., Wabbah, A., Worley, S., Weinstein, G., Weinstein, I. R.,  
Williamson, A., Morley, K., Wren, J., Wrensford, G., Xiang, A. M.,  
Yang, R., Yu, W., Zhou, X., Kucheralapati, R., Nelson, D. and Gibbs, R.

TITLE	JOURNAL	REFERENCE
Direct Submission	Unpublished	2 (bases 1 to 173153)

Worley, K.C.

Worley, K.C.

Direct Submission

Submitted (24-JUL-1999) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 173153)

3 (bases 1 to 173153)

Worley, K.C.

Direct Submission

Submitted (28-SP-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
4 (bases 1 to 173153)

4 (bases 1 to 173153)

Worley, K.C.

### Direct Submission

Submitted (09-OCT-1999) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
5 (bases 1 to 173153)

5 (bases 1 to 173153)

Worley, K.C.

Direct Submission

Submitted (16-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Sep 28, 1999 this sequence version replaced qi:5776564.

On Sep 28, 1999 this sequence version replaced gi:5776564.

INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email

gc-help@bcm.tmc.edu

**CLONE LENGTH:** This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

### ANNOTATION OF FEATURES:

STSSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

**SEQUENCING READ COVERAGE:** Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT-----

----- Summary Statistics

```
Contig length:
Phrap values in estimate:
Average error rate (BCM-Phrap estimate):
Fraction of Phrap values less than 40 :
Number of consensus changing edits:
Number of N's in consensus :
```

----- Consensus changing edit

Position	Original+Context	Edited+Context
8572	caaaaaaa(n)aaaaatagcc	caaaaaaaa(a)aaaaatagcc
23303	gggggtgggg(n)gggtcagggt	gggggtgggg(g)gggtcagggt
35436	aactggctat(n)acatgcagaa	aactggctat(t)acatgcagaa
59324	ttcttttt(n)tttttttttt	ttcttttt(t)tttttttttt
59488	agcctaatt(n)tgatatttt	agcctaatt(t)tgatatttt
59489	ggctaattt(n)tgatatttta	ggctaattt(t)tgatatttta
78551	caagctggt(n)cacacaggg	caagctggt(t)cacacaggg
78561	caacacagg(n)ccatcagggt	caacacagg(c)ccatcagggt
78562	ctggatctga(n)gctcttgta	ctggatctga(t)gctcttgta
78594	ttcatatgc(n)atttcctcac	ttcatatgc(a)atttcctcac
78637	atcatctac(n)ccaaaaggaa	atcatctac(c)ccaaaaggaa
78717	ttgggagcc(n)agatgggtgg	ttgggagcc(g)agatgggtgg
78731	tgggtgatac(n)cnagctcagg	tgggtgata(c)cnagctcagg
78733	gglgatctnc(n)aggtccaggag	gglgatctac(g)aggtccaggag
84931	ctggctttgt(n)ttatnttta	ctggctttgt(t)ttatnttta
84936	ttgtnttat(n)ttttatccta	ttgtnttat(t)ttttatccta
88116	tagaatactg(n)cttctacaat	tagaatact(g)cttctacaat
88153	gggaagaagt(n)ttaatctaat	gggaagaagt(t)ttaatctaat
89301	ggagatagag(n)ccagctctggc	ggagatag(a)ccagctctggc
89333	aaccgcgtct(n)lncTaaaaat	aaccgcgt(c)Tactaaaaat
89335	cccgctctnt(n)ctaaaaatac	cccgctct(c)ctaaaaatac
89474	tcactcagac(n)kgggtgaag	tcactcag(c)kgggtgaag
90120	ttttttttt(g)gggaaaaaaca	ttttttttt(t)gggtagaaca
90123	tttttggg(a)aaaaacagggt	tttttggg(t)agaacacagggt
90127	tttggaaaaa(a)acagggttgt	tttggaaa(a)acagggttgt
102662	tctgaaaaa(n)aaaaatttaata	tctgaaaaa(a)aaaaatttaata
110744	ttccgtctga(n)ggcgggcgca	ttccgtctg(a)ggcgggcgca
110782	gagcccccgc(n)lgcccgcgccg	gagccccc(g)lgcccgcgccg
110866	tgccgactca(n)cgaggcctgt	tgccgactca(g)cgaggcctgt
112388	atcattatga(n)aaatcttagaa	atcattatg(a)aaatcttagaa
126726	ccatccccc(n)acnactcaTca	ccatccccc(c)acnactcaTca
126729	tcaccacac(n)accatcacca	tcaccacac(c)accatcacca
126741	tcacaacac(n)acnactcacca	tcacaacac(c)acnactcacca
126744	tcacaacac(n)accacatTca	tcacaacac(c)accacatTca
126783	ccaccactc(n)ttctccatct	ccaccact(c)ttctccatct
126846	cttctaccat(n)accacacct	cttctaccat(c)accacacct
126963	ttccactgct(n)accactcca	ttccactgt(c)accactcca
145620	caacacaggt(n)acacagcact	caacacagt(c)acacagcact
145676	caatacacag(n)gaacacacac	caatacacag(c)gaacacacac
146199	tcctcccgct(a)ctctcccggt	tcctcccgct(c)ctctcccggt
146210	ctctcccgct(a)ctctccctct	ctctcccgct(c)ctctccctct
155221	ctcccatctt(n)ccccctccct	ctcccatct(c)ccccctccct

----- Distribution of Quality &lt; 40 Bases

[illegible]

Phrap value Range

	3326	AGATGCTAGTAATTACAGAATGGCCCTCTTGTGGGGTGGGCTTGAGGGAAGTCAGCTAT	3388
QY			
	71128	AGATGCTAGTAATTACAGAATGGCCCTCTTGTGGGGTGGGCTTGAGGGAAGTCAGCTAT	71065
Db			
QY	3386	AAACAATTTGCTGGAGCTTTTGTTCAATGGGCTGTGCATTTTTATATATATGTGTTGTAA	3445
Db	71068	AAACAATTTGCTGGAGCTTTTGTTCAATGGGCTGTGCATTTTTATATATATGTGTTGTAA	71009
QY	3446	ATGACATGTCAGGCCCTTGTTTTCATGTTTCCCTAAAAGCAGAATAATTTGCAACATTTGTGTTT	3505
Db	71008	ATGACATGTCAGGCCCTTGTTTTCATGTTTCCCTAAAAGCAGAATAATTTGCAACATTTGTGTTT	70944
QY	3506	GATAGGAATTAATTTGTCGCCACTGCTGTGGACTGTGTTTCTTTGGCCTACTGACTAGTAGAC	3565
Db	70948	GATAGGAATTAATTTGTCGCCACTGCTGTGGACTGTGTTTCTTTGGCCTACTGACTAGTAGAC	70885
QY	3566	CTGTGTTGTCTAAACATGAGTTTCAGCCCTTTGGTCTTTTAATACCATGTCAAATGCA	3625
Db	70888	CTGTGTTGTCTAAACATGAGTTTCAGCCCTTTGGTCTTTTAATACCATGTCAAATGCA	70829
QY	3626	AAC TTC AAT TCT CCC CAT TT TAG CTT TTA TTA AA ACT GAC GTT CT CT T CA AA AC T TC TT GC TG	3685
Db	70828	AAC TTC AAT TCT CCC CAT TT TAG CTT TTA TTA AA ACT GAC GTT CT CT T CA AA AC T TC TT GC TG	70768
QY	3686	AATGGTACTCAGATGTGCATTCACATACAGATGTGTTTGAAGTGGGTGTACTTGCTTT	3745
Db	70768	AATGGTACTCAGATGTGCATTCACATACAGATGTGTTTGAAGTGGGTGTACTTGCTTT	70705
QY	3746	ACCTAATAGATGTGTAATAGAACCTTTTGTAGTCAAA	3783
Db	70708	ACCTAATAGATGTGTAATAGAACCTTTTGTAGTCAAA	70671
RESULT 8			
G26688			
LOCUS	G26688	human STS STS_D63879, sequence tagged site.	STS 02-JUN-1996
DEFINITION	G26688		
ACCESSION	G26688.1	GI:1348920	
VERSION			
KEYWORDS		STS; STS sequence; primer; sequence tagged site.	
SOURCE		Homo sapiens STSs derived from sequences in dbEST and the Unigene collection.	
ORGANISM		Homo sapiens	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS		Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
TITLE		Hudson, T.	
JOURNAL		Whitehead Institute/MIT Center for Genome Research; Physically	
COMMENT		Mapped STS Unpublished (1995)	
		Contact: Thomas Hudson	
		Whitehead Institute/MIT Center for Genome Research	
		Whitehead Institute for Biomedical Research	
		9 Cambridge Center, Cambridge MA 02142 USA	
		Tel: 617 252 1900	
		Fax: 617 252 1902	
		Email: thudson@genome.wi.mit.edu	
		Primer A: TACTTCACCTCGGCCCGG	
		Primer B: ATCTCAGCACACCATCCTG	
		STS size: 221	
		PCR Profile:	
		Presoak:	
		Denaturation:	
		Annealing: 56 degrees C	
		Polymerization:	
		PCR Cycles: 35	
		Thermal Cycler:	
		Protocol:	
		Template: 10 ng	
		Primer: each 5 pm	
		dNTPs: each 4 mM	

Taq Polymerase: 0.025 units/ul  
Total Vol: 20 ul

## Buffer:

MgCl<sub>2</sub>: 1.5 mM  
KCl: 50 mM  
Tris-HCl: 10 mM  
pH: 9.3

Derived from dbEST (genbank accession D63879).

## FEATURES

source Location/Qualifiers  
1..771

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/map="518.1 cR from top of Chr12 linkage group"

32..252

primer\_bind complement(232..252)

BASE COUNT 182 a 168 c 187 g 234 t

## ORIGIN

Query Match 20.3%; Score 769.4; DB 11; Length 771;

Best Local Similarity 99.9%; Pred. No. 8.6e-176;

Matches 770; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
QY 2901 TGAACGGGACGCTGGGAGACAGAAATGCTTACTTCTACTCTGGCCGGCGGACCTCCCA 2960
DB 1 TGAACGGGACGCTGGGAGACAGAAATGCTTACTTCTACTCTGGCCGGCGGACCTCCCA 60
QY 2961 CCACCCAGACATGCTGGGATGGACAGGCTGGTGTCTGCTGCTCGCAACACAGA 3020
DB 61 CCACCCAGACATGCTGGGATGGACAGGCTGGTGTCTGCTGCTCGCAACACAGA 120
QY 3021 TGCTCTCGGCTTTAGACAGAAAGGGAGGGTTCTAAGTCAAGAGCCTTTTCAGTGT 3080
DB 121 TGCTCTCGGCTTTAGACAGAAAGGGAGGGTTCTAAGTCAAGAGCCTTTTCAGTGT 180
QY 3081 CCTCATATTAGGGCAGTGGCAGAAAGTACCACCTCTGCAGGCTGGGCCCGGAGGATGTG 3140
DB 181 CCTCATATTAGGGCAGTGGCAGAAAGTACCACCTCTGCAGGCTGGGCCCGGAGGATGTG 240
QY 3141 GTGCTCTGAGATGTTTGTATCTTAAAGACTGAGGCACAGAGGAAACAGACACAC 3200
DB 241 GTGCTCTGAGATGTTTGTATCTTAAAGACTGAGGCACAGAGGAAACAGACACAC 300
QY 3201 TGTGTTTGGACACAGTCTGCAAAATGTTTCTGGCCAGCTCCGCCCTTTTGTATGAC 3260
DB 301 TGTGTTTGGACACAGTCTGCAAAATGTTTCTGGCCAGCTCCGCCCTTTTGTATGAC 360
QY 3261 ACTTCTCTCCACCTGACACACATGTGCCCGTCATCTTTTAAATTTTAAAGATGAA 3320
DB 361 ACTTCTCTCCACCTGACACACATGTGCCCGTCATCTTTTAAATTTTAAAGATGAA 420
QY 3321 ATGGCAGATGCTAGTAATTCACAGAAATGCCCTCTTGTGGGGTGGGCTGAGGGAGTCA 3380
DB 421 ATGGCAGATGCTAGTAATTCACAGAAATGCCCTCTTGTGGGGTGGGCTGAGGGAGTCA 480
QY 3381 GCTATAAAACATTTGCTGGAGTTTGTTCATAGGGGCTGTCATTTTATATATGTTGT 3440
DB 481 GCTATAAAACATTTGCTGGAGTTTGTTCATAGGGGCTGTCATTTTATATATGTTGT 540
QY 3441 TGTAAATGACATGTGACGCCCTTTGTTTCATGTTTCTTAAAGCAGAAATTTTCAACATTT 3500
DB 541 TGTAAATGACATGTGACGCCCTTTGTTTCATGTTTCTTAAAGCAGAAATTTTCAACATTT 600
QY 3501 GTTTTGTATAGAAATTTTGTGCCACCTGCTGTCGACTGTTTCTTGTCCCTAGTACTA 3560
DB 601 GTTTTGTATAGAAATTTTGTGCCACCTGCTGTCGACTGTTTCTTGTCCCTAGTACTA 660
QY 3561 GTGACCTGTGTTGCTTAAACATGAGTTTCAGGCCCTTTTGTGTTTGTAAATACCATGTCAA 3620
DB 661 GTGACCTGTGTTGCTTAAACATGAGTTTCAGGCCCTTTTGTGTTTGTAAATACCATGTCAA 720
```

QY 3621 ATGCAAACTCAATTCTCCCATTTAGCTTTTAACTGAGCTTCTCTTC 3671  
DB 721 ATGCAAACTCAATTCTCCCATTTAGCTTTTAACTGAGCTTCTCTTC 771

## RESULT 9

AC122224\_3/c

## WPCOMMENT

Sequence split into 4 fragments LOCUS AC122224 Accession AC122224

Fragment Name Begin End

AC122224\_0 1 110000

AC122224\_1 100001 210000

AC122224\_2 200001 310000

AC122224\_3 300001 386096

Continuation (4 of 4) of AC122224 from base 300001 (AC122224 Mus musculus chromosome

Query Match 8.9%; Score 338.2; DB 2; Length 86096;

Best Local Similarity 59.5%; Pred. No. 8e-71;

Matches 1088; Conservative 0; Mismatches 578; Indels 163; Gaps 24;

```
QY 17 GACTGGCGCGAAACCTCGGCTTCAGAACCCGAGGCTGAGTCCAGGCTGGGCCCAAGGC 76
DB 62146 GACAGAGTGCACCCCACTTTGGAGCCAGATCTGAGTCTACGCTCTGGGTACCAAGA 62087
QY 77 TGACGGAGAGGAGGATGAGGTTAAGCGGCTAGGACAAAGAGAAAGTGTATTTCGGCG 136
DB 62086 GCAAAAGAGGAAGATGAATTCGAATGCAGCTGGATGAGAAAGAGTGTGACATCATGG 62027
QY 137 TGTGGCCGCTGGACATACAAACCATGGGGCCAGGCTGGGATCAGCAGAGGAAGGCGT 196
DB 62025 TGTGGCATGAAGG---TACAAGAGCATGGGGCCCAAGATGGAATCAGCAG--TGAAGGCAC 61972
QY 197 GAGCGAGAGCGGATGGGATGAGTAC---GCCATGGCTTCTCCGGGAGAGCTCCCCCGG 253
DB 61971 AAGTGAGAGAAATGGGATGAGGACTGTGCTCTTGACTCTTCTGGGAGGACTCTCTTGG 61912
QY 254 GGAGTACGAGTGGGAAATATGACAAAGAGGAGAGAAACCAGC----- 297
DB 61911 GAAATGCAAGGAATGAAGAGAGCATCTGGAGTAGCCAAAGAATATCTATAAGATCTAT 61852
QY 298 -----TGGAGATTGAGAGACTGGAGGAGGACTGTCTAT 331
DB 61851 GATAACCTAAACCAATGCACCATTTGTTATTAATTCACAGTGAAGAAAGAAATGGAATA 61792
QY 332 CAACGCTCTATGACTACAACCTGGCATGTGACTTGTATCAGACTGCTCAGCTGGAAGGGA 391
DB 61791 GAAATCCCATAACTCCCTTTGTGTGTGACCTGACAGGCTGTGCTGCACTTGAATGAGA 61732
QY 392 GCTTACCAAGTGAG-GATGGCCCGCCAGAAAGATGAGTGAATCTTTCCCTTGACTGAAG 450
DB 61731 GCTGGCTATGGTAAGTGATAGAATCCTATAAGAAGAGACAGTCTCTCCCT---TAAAG 61676
QY 451 AGCTCTGGCTGAGTGGCTGTCATGAGGAGATCAGATGCCCCAGGATGCCCTGGACAGAG 510
DB 61675 AACTCCTGCTAGGCTTACTGTAT-----AGAAGCAAGTCCCTAGGTAGTGTGGCTCTGG 61622
QY 511 AGCAGCTGTATGACCTCTTTGAGAAAGCGCTCAAGGATTTACATTTTCTCTAACTTTGGC 570
DB 61621 ACATATTAACAACTCTCTGTGAGACAGCCATTAAGAGACTACAGCTGTACTGATATCTGTA 61562
QY 571 TAGATATGGCCAGTACTCAGTTGGTGGGATTTGGTCAGAAAGGTGGCCCTTGAGAAAG--- 627
DB 61561 TAGGCTATATACAGACCTTAGTTGTT-GGCTGGGCCAGAAAGTAGCTGGAGAGAGGT 61503
QY 628 -----TTCGCTCCGCTTTTGAAGGCTCT 652
DB 61502 TTTTGTGTTGTTGTTGTTGTTGTTT-----TCTGTGCTTAAAGAGAAAGTTTGT 61443
QY 653 CTCGCTGCTGTTTACATATGACCAAGAGACTCCGCTCTCGGAGGCTTACCAGAGATT 712
DB 61442 GTCATCTCTGATTTTACATATGCCAAGTACAGTCCCATCTGGAGGTTTACTGTGAGAC 61383
QY 713 TGAAGTGCAGTTGTGGAAGCTGCTCGGCTTGAGAAAGTCCACAGTCTCTTTCCGCGGACA 772
```

```
Db 61382 TGAAGTGCCTTACATAGAGCTGCTTAGCTTGAGAAATCTATATCTATGCTCCAGTGACA 61323
Qy 773 GTTGGCATCCACTCTATGATATGAGGCGACATTTGCAGAGTATGAAGATGGTTCAGA 832
Db 61322 TTTGGTATACCACTATGAATG-----AAATTTGCATAGTGTGAAGACTGGTTCAGA 61269
Qy 833 AGACCCAAATACAGAGTCAGTAATTCAGAACTATAACAAGACACTACAGCAGCTGGAGAA 892
Db 61268 AGACCTAAGAGCACATTCAGCCATCCAGAACTATATAA-----CCCTGCAGCAGTTGAA 61214
Qy 893 ATATAAACCCTATGAAGAGCAGCTGTGTGCAGGAGAGGACCAAGGCTGGCAGAAATATCA 952
Db 61213 AAAAAAATACCACTAAGACATAGTTTTCGAAGCTGAGGTTCTGAGGTTGGCAGAAATGCCA 61154
Qy 953 AGCATATATGATTTTTCAGATGAATGCGATCCCTGCTCGCATTCAGTTGATCTTTGA 1012
Db 61153 ACTACATAT---ATTTTGAAGAAATGGATATATCTGCTTGCATTCAGTTGATCTTTGA 61097
Qy 1013 GCGCGCCCTGGTGCAGAACTGCCCTTGCCAGAGCTTATGGATCGGTTACAGTCAGTACCT 1072
Db 61096 GTATGGCTGGTTAAGAAATGCCCTTGGCCACACTTAGAGA---CTACTGACAGTACCT 61041
Qy 1073 AGATCGACAAGTGAAGTAAGGATTTGGTTTATCTGTATCA---TAACCGCGCTATAG 1129
Db 61040 AGATGGAAGCTGAAGTAAGTAATCTCGGTTTCTCTATTACAGCAATGCCACTGGTAG 60981
Qy 1130 AAATGCGCCCTGGAGAGTGGCTTATGGAGTCGTACTCTTGTGCCATGGAGACACATGG 1189
Db 60980 AAATGCTTCCGGACAAAATATTTGTAGATTTGGTATCTCTTGACATTCGTAACAAATGA 60921
Qy 1190 AGTTGATCATCAAGTAATTTCTGTAACTTCGAGAAAGCTTTTGAATCCGCGCTTCATCCA 1249
Db 60920 AGCTGATCATCAGATGACTTTTGTGGTCTTTGGGAAGGCTCTGATGCTGCTTCATCTA 60861
Qy 1250 GGCACACTGATATGTGGAGATTTGGCAGGCATACCTTGATTAACCTGAGGAGAAAGGTTGA 1309
Db 60860 GGCCATTGTATAT---GGGATTTGGCTGGTTTACTTGTATGTCCT-----AAGGGCTG 60811
Qy 1310 TTTCAACAAGACTCCAGTAAAGAGCTGGAGAGTGTAGGCGCCCTTTACTCTGCTCTT 1369
Db 60810 ATTCAACAAGACTATATTAGGAG--AGTTTGGTTATTTATCTACAGCTCTTTTGGGG 60753
Qy 1370 GGAGTATCTGAAGCAGAGGTGGAAGAGCGTTTCAATCAGAGTGGTGTATCCAAAGCTGCGT 1429
Db 60752 TATCCATCTGCTTCAGAAAGTTGAAGGATATTTCAAT---AGTGGAGTTTCAAGGCTCTT 60696
Qy 1430 GATTATGCAGAACTGGGCTAGGATTTAGGCTGAGGCTGCAATTAACATGCAAGAGCTCG 1489
Db 60695 GGTAAAGCAGAAATTTGGATCAGGACTGAGGCTCACTGCACCTTAATGTGCAAGAACCTCA 60636
Qy 1490 GGAACCTCTGGATAGCATATGACATGACAGAGGAAATGCCAAGTACCCCAATGTGGCTAGA 1549
Db 60635 GAACGCGAGGACAGCACATCAGCAAGAACCAAG-----AATGTATAGCTTCA 60587
Qy 1550 GTATTAAACCTGGAAGAGCTCATGTGTACACCCAGCAGCTCCGGAAGGCTCTGCACCG 1609
Db 60586 TTGTTATACACT-GAAAGAAATGCTCAGTGTACACAGAAATTTGCTAG-AGGTTCTACACTG 60529
Qy 1610 GGCCTGCTGACAGTGTACTACCCAGAGCAGCTGTGCAAGTGTGTACTACCATGGA 1669
Db 60528 GGTGGAG-----AGGACTCATCCCGACATGTCTGTGAATCTGTGTACTCAATATGGA 60478
Qy 1670 GAGCAGACAGGTTCTTTTAGAAGATTTGGATATAGCTGTCTCAGAAAACTGAACCCGAT 1729
Db 60477 GAGATAG-AGTCAATAAGAGACTGAGATGTGGTGTCTGAAACTGAAC-----TC 60423
Qy 1730 AGCTGCTGCTCAATGAGCAGAAATGAAG 1758
Db 60422 AACTCATGCTCATGATGTTGAATGAAGG 60394
```

RESULT 10

AC126267/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

AC126267  
Mus musculus chromosome UNK clone R23-191M12, WORKING DRAFT  
SEQUENCE, 14 unordered pieces.  
AC126267  
HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 319737)  
McPherson, J.D. and Waterston, R.H.  
The sequence of Mus musculus clone  
Unpublished  
2 (bases 1 to 319737)  
McPherson, J.D. and Waterston, R.H.  
Direct Submission  
Submitted (04-JUL-2002) Genome Sequencing Center, 4444 Forest Park  
Parkway, St. Louis, MO 63108, USA

----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: http://genome.wustl.edu/gsc/index.shtml  
Contact: submissions@watson.wustl.edu  
----- Project Information -----  
Center project name: M\_BA0191M12  
----- Summary Statistics -----  
Sequencing vector: M13; 0%  
Chemistry: Dye-primer ET; 0% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 309491 bases at least Q40  
Consensus quality: 311769 bases at least Q30  
Consensus quality: 313307 bases at least Q20  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 14 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1316: contig of 1316 bp in length  
\* 1317 1416: gap of unknown length  
\* 1417 3096: contig of 1680 bp in length  
\* 3097 3196: gap of unknown length  
\* 3197 8084: contig of 4888 bp in length  
\* 8085 8184: gap of unknown length  
\* 8185 14192: contig of 6008 bp in length  
\* 14193 14292: gap of unknown length  
\* 14293 18775: contig of 4483 bp in length  
\* 18776 18876: gap of unknown length  
\* 18876 27060: contig of 8185 bp in length  
\* 27061 27160: gap of unknown length  
\* 27161 36165: contig of 9005 bp in length  
\* 36166 36266: gap of unknown length  
\* 36266 42770: contig of 6505 bp in length  
\* 42771 42870: gap of unknown length  
\* 42871 83818: contig of 40948 bp in length  
\* 83819 109449: contig of 25531 bp in length  
\* 109450 109549: gap of unknown length  
\* 109550 176791: contig of 67242 bp in length  
\* 176792 176891: gap of unknown length  
\* 176892 219804: contig of 42913 bp in length  
\* 219805 219904: gap of unknown length  
\* 219905 318952: contig of 99048 bp in length  
\* 318953 319052: gap of unknown length

```
FEATURES * 319053 319737: contig of 685 bp in length.
source Location/Qualifiers
1. .319737
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosomes="UNK"
/clone="RP23-191M12"
1. .1316
/misc_feature /note="assembly_name:Contig13"
1417. .3096
/misc_feature /note="assembly_name:Contig14"
3197. .8084
/misc_feature /note="assembly_name:Contig15"
8185. .14192
/misc_feature /note="assembly_name:Contig16"
14293. .18775
/misc_feature /note="assembly_name:Contig17"
18876. .27060
/misc_feature /note="assembly_name:Contig18"
27161. .36165
/misc_feature /note="assembly_name:Contig19"
36266. .42770
/misc_feature /note="assembly_name:Contig20"
42871. .83818
/misc_feature /note="assembly_name:Contig21"
83919. .109449
/misc_feature /note="assembly_name:Contig22"
109550. .176791
/misc_feature /note="assembly_name:Contig23"
176892. .219804
/misc_feature /note="assembly_name:Contig24"
219905. .318952
/misc_feature /note="assembly_name:Contig25"
319053. .319737
/misc_feature /note="assembly_name:Contig8"
BASE COUNT 91273 a 65131 c 64726 g 97306 t 1301 others
ORIGIN
Query Match 8.9%; Score 338.2; DB 2; Length 319737;
Best Local Similarity 59.5%; Pred. No. 8.7e-71;
Matches 1088; Conservative 0; Mismatches 578; Indels 163; Gaps 24;
QY 17 GACTCGGCGGAAACCTCGGCTTCAGAACCCGAGGCTGAGTCCAAAGGCTGGGCCCAAGGC 76
D 199732 GACAGCAGGTGCAACCCCAACTTTGGAGCCAGACTAGTCTGCTGGGTACAAGA 199673
QY 77 TGACGAGAGGAGGATGAGTTAAGCGCGCTAGGACAGAGGAGGAGGTTATCGCGGC 136
D 199672 GGAAAAAGGAGAGATGAATTGAATGCGAGCTGGATGAGAGGATGTGACATCATGGGC 199613
QY 137 TGTGGCGGCTGGCAGATACAAGACCATGGGCGAGCGTGGGATCAGCAGGAGGAAGCGCT 196
D 199612 TGTGGCATGAGG---TACAAGAGCATGGGCGCAGATGAGATCAGCAG--TGAAGGCAC 199558
QY 197 GAGCGAGAGGATGGGATGAGTAC---GCCATGGCTTCCTCCGCGGAGAGCTCCCGCG 253
D 199557 AAGTGAGAGAATGGGATGAGGAGTGTCTTGTGACTTCTTGGGAGGAGCTCTCTGG 199498
QY 254 GAGTACGAGTGGGAATATGACNAGAGGAGGAGGAACACAGC----- 297
D 199497 GAATGCAAGGAATGAAGAGCATCTGGAGTAGCCAAAGAAATATCTATAAGATCTAT 199438
QY 298 -----TGGAGATTGAGAGACTGGAGGAGCAGTTGTCTAT 331
D 199437 GATAACCTAAACCAATGCACCATCTTTATAAATTCACACGTTGAAAAGAAATGGAATA 199378
QY 332 CAAGCTCTATGACTACAACTGCCATGTCATGATGATGATGATGATGATGATGATGATGAT 391
D 199377 GAAATCCCAATAAATCTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3918
QY 392 GCTTACCAAGGTGAG-GATGGCGCCGCCAAGACATGAGTGAATCTTCCCTTGACTGAG 450
D 199317 GCTGGCTATGGTAAGTGATAGAATCCTATAAGAGAGACAGTCTTCCCTCTTCCCTCTT 199262
```



```

Oy 1490 GGAACCTCTGGATGACATCATGACACGAGGAAATGCCAAGTACCCACATGTGGCTAGA 1549
Db 198221 GAAACGAGGACGACACATCATGACACGAGGAAATGCCAAGTACCCACATGTGGCTAGA 198173
Oy 1550 GTATTACAACCTGAAAGAGCTCATGTGTGACACACCCAGCTGCCGGAAGGCTCTGCACCG 1609
Db 198172 TTGTTATACACT-GAAAGAAATGCTCAGTGTGACACAGAAATGCTAG-AGGTTCTACACTG 198115
Oy 1610 GGCCGCTCAGTGACACCACTGACTACCCAGAGAGCTGTGCGAAGTGTACTTACCCATGGA 1669
Db 198114 GGTGGAG-----AGGACTCATCCCGACATGCTGTGAACTGTACTTCAATATGGA 198064
Oy 1670 GAGACAGAGAGTCTTTTGAAGATGGATATAGCTGTTCAGAAAACCTGAAACCCGATT 1729
Db 198063 GAGAAATAG-AGTCAATAAGAGAGACTGAGATGTGCTGTGAACTGTGAAACCTGAAAC-TC 198009
Oy 1730 AGCTCGTGTCAATGACGACAGAGAAATGAAGG 1758
Db 198008 AACTCATGTGTCATGATGATGAATGAAGG 197980

RESULT 11
AC128917/c
LOCUS
DEFINITION AC128917 94682 bp DNA linear HTG 24-JUL-2002
***, 64 unordered pieces.
ACCESSION AC128917
VERSION AC128917.1 GI:21953478
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Rattus norvegicus.
ORGANISM Rattus norvegicus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 94682)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amarantunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,M.R., David,R.,
Davilla,M.L., Davis,K., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hoques,M., Holloway,C., Hollins,B.,
Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,K., Jolivet,S., Joudah,S.,
Karissone,C., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W., Louissege,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mathewney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogih,M., Okwuonu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rivas,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I.,
Sodergren,E., Sonaite,T., Sparks,A., Stanley,H., Stone,H.,
Tatone,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Taylor,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wlarczyk,R., Wooden,S., Worley,K.,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Unpublished
2 (bases 1 to 94682)
Worley,K.C.
Direct Submission
Submitted (24-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: RBDL
Center clone name: CH230-337D11
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 24222 bases at least Q40
Consensus quality: 26896 bases at least Q30
Consensus quality: 28479 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 64 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1
* 1137: contig of 1136 bp in length
* 1236: gap of unknown length
* 1237: contig of 1240 bp in length
* 2477: gap of unknown length
* 2576: gap of unknown length
* 3660: contig of 1084 bp in length
* 3661: gap of unknown length
* 3761: contig of 1009 bp in length
* 4770: gap of unknown length
* 4870: contig of 1045 bp in length
* 5915: gap of unknown length
* 6013: contig of 1007 bp in length
* 7022: gap of unknown length
* 7122: contig of 1142 bp in length
* 8263: gap of unknown length
* 8264: contig of 1753 bp in length
* 8364: contig of 10116 bp in length
* 10117: gap of unknown length
* 10216: contig of 1027 bp in length
* 11243: gap of unknown length
* 11344: contig of 1161 bp in length
* 12504: gap of unknown length
* 12605: contig of 1282 bp in length
* 13887: gap of unknown length
* 13986: contig of 1186 bp in length
* 15172: gap of unknown length
* 15173: contig of 1050 bp in length
* 15272: gap of unknown length
* 16322: contig of 1096 bp in length
* 16323: gap of unknown length
* 17518: contig of 1096 bp in length
* 17519: gap of unknown length
* 17618: contig of 1142 bp in length
* 18760: gap of unknown length
* 18860: contig of 1046 bp in length
* 19906: gap of unknown length
* 20006: contig of 1039 bp in length
* 21045: gap of unknown length
* 21145: contig of 1025 bp in length
* 22170: gap of unknown length
* 22171: contig of 1173 bp in length
* 23443: contig of 1173 bp in length

```

```
* 23444 23543: gap of unknown length
* 23544 24601: contig of 1058 bp in length
* 24602 24701: gap of unknown length
* 24702 25751: contig of 1050 bp in length
* 25752 25851: gap of unknown length
* 25852 27480: contig of 1629 bp in length
* 27481 29147: gap of unknown length
* 29148 29247: gap of unknown length
* 29248 30745: contig of 1498 bp in length
* 30746 30845: gap of unknown length
* 30846 31872: contig of 1027 bp in length
* 31873 33109: contig of 1137 bp in length
* 33110 33209: gap of unknown length
* 33210 34648: contig of 1439 bp in length
* 34649 34748: gap of unknown length
* 34749 36227: contig of 1479 bp in length
* 36228 36327: gap of unknown length
* 36328 37444: contig of 1117 bp in length
* 37445 37544: gap of unknown length
* 37545 38867: contig of 1323 bp in length
* 38868 38967: gap of unknown length
* 38968 40000: contig of 1033 bp in length
* 40001 41996: contig of 1895 bp in length
* 41996 42095: gap of unknown length
* 42096 43651: contig of 1556 bp in length
* 43652 43752: gap of unknown length
* 43752 45104: contig of 1352 bp in length
* 45104 45204: gap of unknown length
* 45204 46389: contig of 1185 bp in length
* 46389 46489: gap of unknown length
* 46489 48184: contig of 1596 bp in length
* 48185 50250: contig of 2066 bp in length
* 50251 51366: gap of unknown length
* 51367 51467: gap of unknown length
* 51467 53345: contig of 1878 bp in length
* 53345 53445: gap of unknown length
* 53445 54818: contig of 1373 bp in length
* 54818 54918: gap of unknown length
* 54918 56537: contig of 1620 bp in length
* 56538 58436: contig of 1799 bp in length
* 58437 58537: gap of unknown length
* 58537 59827: contig of 1290 bp in length
* 59827 59927: gap of unknown length
* 59927 61038: contig of 1111 bp in length
* 61038 62247: gap of unknown length
* 62247 62347: gap of unknown length
* 62348 63507: contig of 1160 bp in length
* 63508 65100: gap of unknown length
* 65101 65200: gap of unknown length
* 65201 66425: contig of 1224 bp in length
* 66425 66525: gap of unknown length
* 66525 68133: contig of 1608 bp in length
* 68133 68233: gap of unknown length
* 68233 70195: contig of 1962 bp in length
* 70195 70295: gap of unknown length
* 70295 71985: contig of 1690 bp in length
* 71985 72085: gap of unknown length
* 72085 73711: contig of 1626 bp in length
* 73711 75626: contig of 1815 bp in length
* 75626 76737: gap of unknown length
* 76737 78863: gap of unknown length
* 78863 78963: contig of 2027 bp in length
* 78963 78963: gap of unknown length

* 78964 80160: contig of 1197 bp in length
* 80161 80260: gap of unknown length
* 80261 81636: contig of 1376 bp in length
* 81637 81736: gap of unknown length

Query Match: 8.1%; Score 306.2; DB 2; Length 94682;
Best Local Similarity 80.3%; Pred. No. 4.8e-63;
Matches 372; Conservative 0; Mismatches 88; Indels 3; Gaps 1;

QY 1923 GAAGACAGCCCTCCAAACGACAGAGGGTCGAGAACAGCATCCCTGCAGCTGGAGAACA 1982
DB 71480 GCAGAACAGCCCTCCAAACGACAGAGGATGGAGAACAGCTTG---GCCTCTGGAGTGGCT 71424
QY 1983 CAAATGTAGAGTAGCAGCAGGCGCGCTGGGAAATGTCTGCCGTAGATGGAGGCC 2042
DB 71423 TCAGCTAAGGAGGAGAACAAACATTTCCGGGAAATGCTTACATATAGACGTGCTCCT 71364
QY 2043 CCTTCGAGCAGAGGAGGAGGAGCGCTCCCTCGAAGAGGACATGCCAAGGTGCTGCAC 2102
DB 71363 CCTTCCAGCAGAGGAGGAGGAGCGCGCTCCCTTAAGCGGACATGCCAAGGTGCTGCAC 71304
QY 2103 GACAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2162
DB 71303 GACAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 71244
QY 2163 CGGACACCAAGCTCAGGCGCACTCTTCGAGGCGCTGTGGGAGGTGTCAGATCCGACCC 2222
DB 71243 CCAGAGGTGAAGCTCAGGCGCGCTCTTTGAGGCGCTGTGGGAGGTGTCAGATCCGAC 71184
QY 2223 ATCTTCAGCAACCGTGGGAGGATTTCCGAGGTTACTGCTAGCTGGAGGTTTAAAGAGAG 2282
DB 71183 ATCTTCAGCAACCGTGGGAGGATTTCCGAGGTTACTGCTAGCTGGAGGTTTAAAGAGAG 71124
QY 2283 TCAGCCCTTCAGCAGCTGAGATGGACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2342
DB 71123 TCAGCCCTTCAGCAGCTGAGATGGACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 71064
QY 2343 TCCCTCTGTGTGGTAAAGACAAACCCCGCATTTTAAAGGTGT 2385
DB 71063 TCCCTCTGTGTGGTAAAGACAAACCCCGCATTTTAAAGGTAT 71021

RESULT 12
AX312274
LOCUS AX312274
DEFINITION Sequence 5259 from Patent WO0190366.
ACCESSION AX312274
VERSION AX312274.1 GI:17897367
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Leach, M.D. and Shimkets, R.A.
TITLE Human polynucleotides and polypeptides encoded thereby
JOURNAL Patent: WO 0190366-A 5259 29-NOV-2001;
Curagen Corporation (US)
FEATURES
Location/Qualifiers
source
1..324
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 70 a 82 c 104 g 68 t
ORIGIN

Query Match: 6.1%; Score 233.4; DB 6; Length 324;
Best Local Similarity 82.7%; Pred. No. 1.6e-45;
Matches 267; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 1110 GTACATAACCGCGCTATTAGAACTGCCCTCGACAGTTTGCCTTATGGAGTCGGTACCTC 1169
DB 1 GTACACAGCGCTGCTGCGGAACTGCCCATGGACAGTTGCCCTGTGGAGTCGGTACCTT 60
```

QY 1170 TTGCCATGGAGACATGGATTGATCATCAAGTAATTTCTGTAACTTCGAGAAAGCT 1229  
 Db 61 CTGCCATGGAGACATGGATTGATCATCAAGTAATTTCTGTAACTTCGAGAAAGCT 120  
 QY 1230 TTGAATCCGCTTCATCCAGGCCACATGATATGTGTGAGATTGGCAGGCATACCTTGAT 1289  
 Db 121 CTGAGTCCGCTTCATCCAGGCCACATGATATGTGTGAGATTGGCAGGCATACCTTGAT 180  
 QY 1290 TACCTGAGGAGAGGGTTGATTTCAAAAGAGTCCAGTAAAGAGCTGGAGGAGTTCAGG 1349  
 Db 181 TACCTGAGGAGAGGGTTGATTTCAAAAGAGTCCAGTAAAGAGCTGGAGGAGTTCAGG 240  
 QY 1350 GCCCCTTACTCTGCTTGGAGTATCTGAAGCAGGAGTGGAGAGCGCTTCAATGAG 1409  
 Db 241 TCCATGTTTCAGCGAGCTCTGGAGTACCTCAGCAGGAGGTTGAGGAGCGCTTTCAGCGAG 300  
 QY 1410 AGTGGTATCCAGCTCGTGTAT 1432  
 Db 301 AGTGGGATCCAGCTCGTGTAT 323

## RESULT 13

G62164

LOCUS

DEFINITION EST135 Foetal cDNA library directionally cloned in Unizap XR  
 vector Equus caballus STS CDNA 3', sequence tagged site.

ACCESSION

G62164

VERSION

G62164.1

KEYWORDS

STS.

SOURCE

Equus caballus.

ORGANISM

Equus caballus

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

314 bp mRNA linear STS 01-MAY-2000  
 G62164  
 EST135 Foetal cDNA library directionally cloned in Unizap XR  
 vector Equus caballus STS CDNA 3', sequence tagged site.  
 G62164.1 GI:7677606  
 STS.  
 Equus caballus.  
 Equus caballus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Perissodactyla; Equidae; Equus.  
 1 (bases 1 to 314)  
 Brandon, R., Thomas, S. and Bell, K.  
 Development of equine expressed sequence tags for use in radiation  
 hybrid mapping  
 Unpublished

Contact: Richard Brandon  
 Australian Equine Blood Typing Research Laboratory  
 University of Queensland  
 St Lucia, Brisbane, Queensland, 4072, Australia  
 Tel: 61733658501  
 Fax: 61733654899  
 Email: brando@plp.uq.edu.au  
 Primer A: CTCACACTGTTTATAGCAATCTG  
 Primer B: CAAATACATCCATATTGAATGC  
 STS size: 241  
 PCR Profile:

Cycle 1:  
 95 degrees C for 3.00 minute  
 60 degrees C for 1.00 minute  
 72 degrees C for 1.00 minute  
 Cycles 2-31:  
 95 degrees C for 0.45 minute  
 60 degrees C for 1.00 minute  
 72 degrees C for 1.00 minute  
 Cycle 32:  
 95 degrees C for 0.45 minute  
 60 degrees C for 1.00 minute  
 72 degrees C for 10.00 minute.

Thermal Cycler: MJ Research PTC-100

Protocol:

Template:

each 0.5 uM

dNTPs: each 200 uM

Taq Polymerase: 0.02 units/ul

Annealing Temp: 60

Total Vol: 20 ul

Buffer:

MgCl2: 1.5 mM

(NH4)2SO4: 16.6 mM

Tris-HCl: 67 mM

Gelatin: 200 ug/ml

Triton X-100 0.45%

PCR works in the horse and not mouse for specific use in RH panel  
 screening

## FEATURES

source

location/Qualifiers

1..314

/organism="Equus caballus"

/db\_xref="taxon:9796"

/tissue\_type="60-day foetus"

/clone\_lib="Foetal cDNA library directionally cloned in

Unizap XR vector"

/note="vector: Bluescript; V-type: phagemid; A whole  
 foetus was used as a source of RNA in the construction of a  
 Strategene Unizap XR directional library. The phagemid  
 Bluescript was excised from the virus vector and used as  
 template for sequencing reactions from the 3' end of the  
 cDNA."

gene

1..314

/gene="SART1"

/note="Squamous cell carcinoma antigen (SART-3)"

STS

1..241

/gene="SART1"

1..26

/gene="SART1"

complement(218..241)

primer\_bind 92 a 49 c 119 t

BASE COUNT 54 g 119 t

ORIGIN

Query Match 5.1%; Score 194; DB 11; Length 314;

Best Local Similarity 84.0%; Pred. No. 6e-36;

Matches 257; Conservative 0; Mismatches 40; Indels 9; Gaps 3;

QY 3502 TTTTGTATAGAAATATTGTCACCATCTGCTGAGTCTTCTTCCCTAGTACGAC 3561

Db 9 TGTATTAGCAATCTGCTGTCACCATCTGCTGAGTCTTCTTCCCTAGTACGAC 68

QY 3562 TGACC-----TGTGTGTCTAAACATGAGTTTCAGCCCTTGGTTTGTGTTT-AAATACCATG 3616

Db 69 TGAACCTTGTCTTCTTGTCTAAATATTGAGTTTCAGCCCTTGGTTTGTGTTT-AAATACCATG 128

QY 3617 TCAATGCAAACTTCAATTCCTCCCATTTAGC---TTTATTAACCTGACGCTTCTCTCA 3672

Db 129 TCAAGTGCAAACTTAAATTTGCCCATTTAGCTTTGTTTATTAACCTGAGTCTCTCTCA 188

QY 3673 AAACCTCTTGTCTGAATGCTACTCAGATGTCATTCACATACAGATGTTTGAAGTGGG 3732

Db 189 AAAATCTTCCGGAATGGTACTCAGATGTCATTCACATACAGATGTTTGAAGTGGG 248

QY 3733 TGTACCTTGCCTTTACCTAAATAGATGTGTAAATAGAACTTTTGTAAAGTCAAAAAA 3792

Db 249 TGTACCTTGCCTTTCTCTAAATAGATGTGTAAATAGAACTTTTGTAAAGTCAAAAAA 308

QY 3793 AAAAAA 3798

Db 309 AAAAAA 314

## RESULT 14

HS53G8F

LOCUS

DEFINITION

H.sapiens CpG island DNA genomic Msel fragment, clone 53g8, forward

read cp953g8.ftib.

ACCESSION

255627

VERSION

255627.1

KEYWORDS

CpG island; genomic Msel fragment.

SOURCE

Homo sapiens.

ORGANISM

Homo sapiens

REFERENCE

AUTHORS

TITLE

JOURNAL

218 bp DNA linear PRI 17-OCT-1995  
 H.sapiens CpG island DNA genomic Msel fragment, clone 53g8, forward  
 read cp953g8.ftib.  
 255627  
 255627.1 GI:1021668  
 CpG island; genomic Msel fragment.  
 Homo sapiens.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 218)  
 Macdonald, M., Huckle, E., Wilkinson, P. and Micklem, G.  
 Direct Submission  
 Submitted (16-OCT-1995) The Sanger Centre, Hinxton, Cambridgeshire.

CB10 1RQ, England. E-mail contact: humquery@sanger.ac.uk  
2 (bases 1 to 218)  
Cross, S.H., Charlton, J.A., Nan, X. and Bird, A.P.  
Purification of CpG islands using a methylated DNA binding column  
Nat. Genet. 6 (3), 236-244 (1994)  
94282070  
MEDLINE  
8012384  
PUBMED  
Vector: pGEM-5zf(-)  
Clones are available from the UK MRC Human Genome Mapping Project  
Resource Centre, Hinxton, Cambridgeshire CB10 1RQ, UK. See URL:  
http://www.hgmp.mrc.ac.uk/ for details  
or contact: biohelp@hgmp.mrc.ac.uk.  
Location/Qualifiers  
1. .218  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="53g8"  
/sex="male"  
/tissue\_type="blood"  
/clone\_lib="CGI-1"  
/dev\_stage="adult"  
BASE COUNT 57 a 43 c 85 g 31 t 2 others  
ORIGIN

Query Match 4.9%; Score 186.8; DB 9; Length 218;  
Best Local Similarity 96.8%; Pred. No. 3.3e-34;  
Matches 211; Conservative 0; Mismatches 4; Indels 3; Gaps 2;  
QY 98 TAAGCGGCTAGGACAGG--AGAAAGGTCTTATCCGGGCTGTGGCCGCTGGGACATAC 155  
|||||  
Db 1 TAAGCGGCTAGGACAGGAGTAGTAAAGGTGTATCCGGGCTGTGGCCGCTGGGACATAC 60  
QY 156 AAGACCATGGCGGCGGCGGATCAGCAGGAGGAGGCGGTGAGCGAGAGCGATGGGGAT 215  
|||||  
Db 61 AAGACCATGGCGGCGGCGGATCAGCAGGAGGAGGCGGTGAGCGAGAGCGATGGGGAT 120  
QY 216 GAGTACGCATGGCTTCCTCCGGGAGAGCTCCCGGGGAGTACAGTGGGAGATGAC 275  
|||||  
Db 121 GAGTACGCATGGCTTCCTCCGGGAGAGCTCCCGGGGAGTACAGTGGGAGTACAGT 180  
QY 276 GAA-GAGGAGGAG 312  
|||||  
Db 181 GAGGAGGAG 218

RESULT 15  
G54201  
LOCUS  
DEFINITION B13G14/T7 Human Chromosome 12 Homo sapiens STS genomic clone  
RPCI-11-B13G14 T7, sequence tagged site.  
ACCESSION G54201  
VERSION G54201.1 GI:5690177  
KEYWORDS STS.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 369)  
Montgomery, K.T., Marondel, I., Kneitz, S., Yang, R., Lau, S., Lee, E.,  
Miller, A., Issar, A., Shim, C., Decker, J., Chiu, D. and  
Kuchelapati, R.  
A Sequence Ready Map of Human Chromosome 12 (1999)  
Unpublished (1999)  
Contact: Raju Kuchelapati  
Kuchelapati Lab., Department of Molecular Genetics  
Albert Einstein College of Medicine (AECOM)  
1300 Morris Park Avenue, Bronx, NY 10461, USA  
Tel: 7184302824  
Fax: 7184308778  
Email: Kuchelra@aecom.yu.edu  
Primer A: AGCGTCCACGTTCTC  
Primer B: GCTCTTTCCGTTGTTACG

STS size: 201  
Protocol:  
template 30-100ng  
each primer 50ng  
dNTP 200mM  
TAQ Gold 0.6U  
MgCl2 2.5mM  
BSA 50mg/ml  
Final Volume 15ml  
Preheat: 10 min 94 deg C  
Denaturation: 20 sec 94 deg C  
Annealing: 20 sec 56 deg C  
Polymerization: 20 sec 72 deg C  
PCR Cycles: 36  
Buffer:  
KCl: 50 mM  
Tris-HCl: 10 mM  
pH: 8.3  
BAC B13G14 from RPCI-11 human BAC Library  
(http://bacpac.med.buffalo.edu/) Maps between AFW304wg5 and  
AFW116xb8.  
Location/Qualifiers  
1. .369  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="RPCI-11-B13G14"  
/clone\_lib="Human Chromosome 12"  
32. .232  
primer\_bind 32. .51  
primer\_bind complement(213. .232)  
BASE COUNT 117 a 73 c 94 g 85 t  
ORIGIN

Query Match 4.3%; Score 162; DB 11; Length 369;  
Best Local Similarity 87.6%; Pred. No. 3.6e-28;  
Matches 177; Conservative 0; Mismatches 25; Indels 0; Gaps 0;  
QY 1756 AGGCTCAG 1815  
|||||  
Db 166 AGGCTCAG 225  
QY 1816 AAAGAGCTCGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1875  
|||||  
Db 226 AAAGAGCTCGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 285  
QY 1876 AGCGCGGACAGATGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1935  
|||||  
Db 286 AGCGCGGACATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 345  
QY 1936 CCAAGCGCAG 1957  
Db 345 TCTTGAG 367

Search completed: June 22, 2003, 08:55:45  
Job time : 6406 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 18, 2003, 13:47:23 ; Search time 46 Seconds  
(without alignments)  
4313.552 Million cell updates/sec

Title: US-09-763-985A-2

Perfect score: 4994

Sequence: 1 MATAAETSAPEAESKAGP.....AATEAPKMSNADFAKLEFLRK 963

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: SPTREMBL\_21.\*
- 2: sp\_archaea.\*
- 3: sp\_bacteria.\*
- 4: sp\_fungi.\*
- 5: sp\_human.\*
- 6: sp\_invertebrate.\*
- 7: sp\_mammal.\*
- 8: sp\_mhc.\*
- 9: sp\_organelle.\*
- 10: sp\_phase.\*
- 11: sp\_plant.\*
- 12: sp\_rodent.\*
- 13: sp\_virus.\*
- 14: sp\_vertebrate.\*
- 15: sp\_unclassified.\*
- 16: sp\_rvirus.\*
- 17: sp\_bacteriap.\*
- 18: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4994	100.0	963	4 Q15020	Q15020 homo sapien
2	4379.5	87.7	962	11 Q9JL18	Q9JL18 mus musculus
3	1812	36.3	364	4 Q86J95	Q86J95 homo sapien
4	902	18.1	816	10 Q8S919	Q8S919 arabidopsis
5	864	17.3	836	5 Q17430	Q17430 caenorhabdi
6	823	16.5	768	10 Q9STW7	Q9STW7 arabidopsis
7	506.5	10.1	1014	3 Q9USY2	Q9USY2 schizosacch
8	425	8.5	941	5 Q9W4D2	Q9W4D2 drosophila
9	420	8.4	943	5 Q18352	Q18352 drosophila
10	349	7.0	826	3 Q01491	Q01491 ophiostoma
11	293.5	5.9	705	13 Q06459	Q06459 xenopus lae
12	280	5.6	744	5 Q16376	Q16376 caenorhabdi
13	272	5.4	1305	5 Q9NFE2	Q9NFE2 drosophila
14	263.5	5.3	665	10 Q9FJ37	Q9FJ37 arabidopsis
15	252.5	5.1	715	11 Q9QZX1	Q9QZX1 rattus norv
16	249.5	5.0	707	11 Q99K50	Q99K50 mus musculus

17	248.5	5.0	610	10 Q9LIH8	Q9LIH8 arabidopsis
18	243	4.9	557	10 Q9FVQ1	Q9FVQ1 arabidopsis
19	241.5	4.8	1009	5 Q9VAT3	Q9VAT3 drosophila
20	238.5	4.8	960	4 Q8BPV6	Q8BPV6 homo sapien
21	236.5	4.7	673	10 Q9FNM3	Q9FNM3 arabidopsis
22	234	4.7	836	4 Q9NYD8	Q9NYD8 homo sapien
23	234	4.7	848	4 Q9BZJ1	Q9BZJ1 homo sapien
24	233.5	4.7	476	5 Q27199	Q27199 tetrahymena
25	232	4.6	690	11 Q9CQC1	Q9CQC1 mus musculus
26	231.5	4.6	724	3 Q9HF03	Q9HF03 cryptococcu
27	229.5	4.6	675	10 Q9LKS1	Q9LKS1 arabidopsis
28	227.5	4.6	571	3 Q8X003	Q8X003 neurospora
29	226	4.5	575	3 Q9Y7A8	Q9Y7A8 neurospora
30	226	4.5	687	4 Q8WY15	Q8WY15 homo sapien
31	225	4.5	687	4 Q9BZJ2	Q9BZJ2 homo sapien
32	224	4.5	836	4 Q9NQH5	Q9NQH5 homo sapien
33	222.5	4.5	599	10 Q8SD35	Q8SD35 arabidopsis
34	221.5	4.4	711	10 Q8S5Y7	Q8S5Y7 oryza sativ
35	221.5	4.4	1972	11 Q8R384	Q8R384 mus musculus
36	219.5	4.4	717	11 Q9RL17	Q9RL17 mus musculus
37	217	4.3	564	4 Q9GZW7	Q9GZW7 homo sapien
38	216.5	4.3	708	4 Q9GQD6	Q9GQD6 homo sapien
39	216.5	4.3	717	4 Q12996	Q12996 homo sapien
40	215	4.3	733	3 Q14233	Q14233 schizosacch
41	214.5	4.3	611	10 Q41042	Q41042 pisum sativ
42	212.5	4.3	524	4 Q14498	Q14498 homo sapien
43	212.5	4.3	594	5 Q9VM49	Q9VM49 drosophila
44	212.5	4.3	883	5 Q9V6S4	Q9V6S4 drosophila
45	211	4.2	495	10 Q9ASP6	Q9ASP6 arabidopsis

## ALIGNMENTS

### RESULT 1

Q15020 PRELIMINARY; PRT; 963 AA.

ID Q15020;  
AC Q15020;  
DT 01-NOV-1996 (TREMREL. 01, Created)  
DT 01-NOV-1996 (TREMREL. 01, Last sequence update)  
DT 01-JUN-2002 (TREMREL. 21, Last annotation update)  
DE ORF.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=96127530; PubMed=8590280;

RA Nagase T., Seki N., Tanaka A., Ishikawa K., Nomura N.;

RT "Prediction of the coding sequences of unidentified human genes. IV.

RT The coding sequences of 40 new genes (K1AA0121-K1AA0160) deduced by

analysis of cDNA clones from human cell line KG-1.";

RL DNA Res. 2:167-174(1995).

RN [2]

RP SEQUENCE FROM N.A.

RA Itoh K., Yang D., Sasatomi T., Nakao M., Shichiho S., Takasu H.,

RA Matsumoto H., Mori K., Yamana H.;

RT "SART-3 (Squamous cell carcinoma antigen recognized by T cells).";

RL Submitted (DEC-1998) to the EMBL/Genbank/DBJ databases.

DR EMBL; D63879; BAA09929.1; -

DR EMBL; AB020880; BAA78384.1; -

DR HSSP; P09012; 2UIA.

DR InterPro; IPR003107; HAT.

DR InterPro; IPR000504; RNA\_rec\_mot.

DR Pfam; PF00076; rrm; 2.

DR SMART; SM00386; HAT; 6.

DR SMART; SM00386; RRM; 2.

DR PROSITE; PS0102; RRM; 2.

DR PROSITE; PS00030; RRM\_RNP\_1; UNKNOWN\_2.

SQ SEQUENCE 963 AA; 109934 MW; 06B26CEB8B819102A CRC64;

Query Match 100.0%; Score 4994; DB 4; Length 963;

Best Local Similarity 100.0%; Pred. No. 2.1e-291;  
Matches 963; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 MATAAETSASEPEAEKAGPKADGDEDEKAAATRRKVLRSRAVAATAATYKTMGPWADQOE 60
DB 1 MATAAETSASEPEAEKAGPKADGDEDEKAAATRRKVLRSRAVAATAATYKTMGPWADQOE 60
QY 61 GVSDEGDEYAMASSAESPGYEWYDEEKEKQLEIERLEQSLINVDYNCHVDLIR 120
DB 61 GVSDEGDEYAMASSAESPGYEWYDEEKEKQLEIERLEQSLINVDYNCHVDLIR 120
QY 121 LRLEGELTKVRMAROKMSEIPLTEELWLEWHDSEISMAQDGLDREHYDIFKAVKYD 180
DB 121 LRLEGELTKVRMAROKMSEIPLTEELWLEWHDSEISMAQDGLDREHYDIFKAVKYD 180
QY 181 ICPNIWLEYGQSVGGIGKGLKVRVFERALSSVGLHMTKGLALWEAYREFESAIVE 240
DB 181 ICPNIWLEYGQSVGGIGKGLKVRVFERALSSVGLHMTKGLALWEAYREFESAIVE 240
QY 241 AARLEKVSFLRRQLAIPLYDMEATFAEYEWSEDPPIESVIQNYNKALQLEKYPYEE 300
DB 241 AARLEKVSFLRRQLAIPLYDMEATFAEYEWSEDPPIESVIQNYNKALQLEKYPYEE 300
QY 301 ALLOAEAPRLAEYQAYIDFEMKIGDPARTOLIIFERALVENCPLVDLWIRYSQYLDRLQKV 360
DB 301 ALLOAEAPRLAEYQAYIDFEMKIGDPARTOLIIFERALVENCPLVDLWIRYSQYLDRLQKV 360
QY 361 KDLVLSVHNRAIRNCPTWVALSRVLLAMERGVHDHOVTSVTFEKNALNAGFTQATDYVEI 420
DB 361 KDLVLSVHNRAIRNCPTWVALSRVLLAMERGVHDHOVTSVTFEKNALNAGFTQATDYVEI 420
QY 421 WQAYLDYLRRRVDFKODSKELEELRAAFTRALEYKQVEERFNEGDPSCVIMONWAR 480
DB 421 WQAYLDYLRRRVDFKODSKELEELRAAFTRALEYKQVEERFNEGDPSCVIMONWAR 480
QY 481 IEARLCNNMOKARELWDSIMTRGNAKYANMWLEYNLERAHGDTQHCRAKALHRAVQCTSD 540
DB 481 IEARLCNNMOKARELWDSIMTRGNAKYANMWLEYNLERAHGDTQHCRAKALHRAVQCTSD 540
QY 541 YPEHVCEVLLTWERTEGSLEDWDIAVQKTETRLARVNEQRMKAKEAALVQOEERKAEQ 600
DB 541 YPEHVCEVLLTWERTEGSLEDWDIAVQKTETRLARVNEQRMKAKEAALVQOEERKAEQ 600
QY 601 RKRAREKALKKKKIRGPEKRGADDEDEKWDGDEDEEQQPSKRRRVENSIPAAGETQNV 660
DB 601 RKRAREKALKKKKIRGPEKRGADDEDEKWDGDEDEEQQPSKRRRVENSIPAAGETQNV 660
QY 661 EVAAGPAGKCAAVDVEPPSKQKKAASLRDMPKVLHDSKSDITVFNLPYSMQEPDT 720
DB 661 EVAAGPAGKCAAVDVEPPSKQKKAASLRDMPKVLHDSKSDITVFNLPYSMQEPDT 720
QY 721 KLRLPFEACGEVYQIRPIFENRGRDFRGYCYVEFKKESALQALEMDRKSVEGRPMFVSPC 780
DB 721 KLRLPFEACGEVYQIRPIFENRGRDFRGYCYVEFKKESALQALEMDRKSVEGRPMFVSPC 780
QY 781 VDKSKNPDKFVRYSTSLKHKLFISGLPFSCTKEELEICKAHGTVDKDLRLVTNRAGRP 840
DB 781 VDKSKNPDKFVRYSTSLKHKLFISGLPFSCTKEELEICKAHGTVDKDLRLVTNRAGRP 840
QY 841 KGLAYVEYENESQASQAVMKMGMTIKENI KVALSNPPQKVPKPEKTRKAPGPMLLP 900
DB 841 KGLAYVEYENESQASQAVMKMGMTIKENI KVALSNPPQKVPKPEKTRKAPGPMLLP 900
QY 901 QTYGARGKRGRTQLSLLPRALQPSAAAPQAEANGPAAAPAAVAPAAEPKMSNADFAPKLF 960
DB 901 QTYGARGKRGRTQLSLLPRALQPSAAAPQAEANGPAAAPAAVAPAAEPKMSNADFAPKLF 960
QY 961 LRR 963
DB 961 LRR 963
```

RESULT 2

Q9JLI8

```
ID Q9JLI8 PRELIMINARY; PRT; 962 AA.
AC Q9JLI8;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Tumor-rejection antigen SART3.
GN SART3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=BALB/C; TISSUE=SQUAMOUS CELL CARCINOMA;
RX MEDLINE=20222349; PubMed=10761712;
RA Harada K., Yamada A., Mine T., Kawagoe N., Takasu H., Itoh K.;
RT "Mouse homolog of the human SART3 gene encoding tumor-rejection
RT antigen.";
RL Jpn. J. Cancer Res. 91:239-247(2000).
DR EMBL; AF172722; AAF65228.1; -.
DR HSSP; P09012; 2U1A.
DR MGD; MGI:1858230; Sart3.
DR InterPro; IPR003107; HAT.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rtm; 2.
DR SMART; SM00386; HAT; 6.
DR SMART; SM00360; RRM; 2.
DR PROSITE; PS0102; RRM; 2.
DR PROSITE; PS00030; RRM_RNP_1; UNKNOWN_2.
SQ SEQUENCE 962 AA; 109618 MW; 23BC235125E7A09C CRC64;
```

Query Match

Best Local Similarity 87.7%; Score 4379.5; DB 11; Length 963;

Matches 839; Conservative. 52; Mismatches 70; Indels 3; Gaps 3;

```
QY 1 MATAAETSASEPEAEKAGPKADGDEDEKAAATRRKVLRSRAVAATAATYKTMGPWADQOE 60
DB 1 MATAAETSASEPEAEKAGPKADGDEDEKAAATRRKVLRSRAVAATAATYKTMGPWADQOE 60
QY 61 GVSDEGDEY-AMASSAESPGYEWYDEEKEKQLEIERLEQSLINVDYNCHVDLIR 119
DB 61 GASGSDGDEDEAMASSAESGAGEDWEYDEEKEKQLEIERLEQSLINVDYNCHVDLIR 120
QY 120 RLRLLEGELTKVRMAROKMSEIPLTEELWLEWHDSEISMAQDGLDREHYDIFKAVKD 179
DB 121 RLRLLEGELTKVRMAROKMSEIPLTEELWLEWHDSEISMAQDGLDREHYDIFKAVKD 180
QY 180 YICPNWLEYGQSVGGIGKGLKVRVFERALSSVGLHMTKGLALWEAYREFESAIV 239
DB 181 YICPNWLEYGQSVGGIGKGLKVRVFERALSSVGLHMTKGLALWEAYREFESAIV 240
QY 240 EAARLEKVSFLRRQLAIPLYDMEATFAEYEWSEDPPIESVIQNYNKALQLEKYPYEE 299
DB 241 EAARLEKVSFLRRQLAIPLYDMEATFAEYEWSEDPPIESVIQNYNKALQLEKYPYEE 300
QY 300 EALLQAEAPRLAEYQAYIDFEMKIGDPARTOLIIFERALVENCPLVDLWIRYSQYLDRLQK 359
DB 301 EALLQAEAPRLAEYQAYIDFEMKIGDPARTOLIIFERALVENCPLVDLWIRYSQYLDRLQK 360
QY 360 VKDLVLSVHNRAIRNCPTWVALSRVLLAMERGVHDHOVTSVTFEKNALNAGFTQATDYVE 419
DB 361 VKDLVLSVHSRAVRNCPTWVALSRVLLAMERGVHDHOVTSVTFEKNALNAGFTQATDYVE 420
QY 420 IQQAYLDYLRRRVDFKODSKELEELRAAFTRALEYKQVEERFNEGDPSCVIMONWA 479
DB 421 IQQAYLDYLRRRVDFKODSKELEELRAAFTRALEYKQVEERFNEGDPSCVIMONWA 480
QY 480 RIEARLCNNMOKARELWDSIMTRGNAKYANMWLEYNLERAHGDTQHCRAKALHRAVQCTSD 539
DB 481 RVEARLCNNMOKARELWDSIMTRGNAKYANMWLEYNLERAHGDTQHCRAKALHRAVQCTSD 540
QY 540 DYPEHVCEVLLTWERTEGSLEDWDIAVQKTETRLARVNEQRMKAKEAALVQOEERKAE 599
```

```
Db 541 DYPEHVCVLLTMEETGLEDWDLAIQKTETRLARVNEQRMKAKEAALVQOEKAE 600
Qy 600 ORKRAKAKKALKKKKIRGPEKRCAGDEDEKWKGDDEEOPSKRRRVENSIPRAGETQN 659
Db 601 ORKVRABKALKKKKIRGADKKRDEDEENGEDEEOPSKRRRVENSIL-ASGEASA 659
Qy 660 VEVAAGPAGKCAAVDVEPPSKOKEKAASLKRDMPKVLHDSKSDITTFVSNLPSMQEPD 719
Db 660 MKEETELSGKCLTIDVGPSPKOKEKAASLKRDMPKVAHDSKSDITTFVSNLPSIEPE 719
Qy 720 TKLRPLFACCEGVQIRIFNRRGDFRGYCVVEKEEKSALQALEMDRKSVEGRPMFVSP 779
Db 720 VKLRPLFVCEGVQIRIFNRRGDFRGYCVVEKEEKSALQALEMDRKSVEGRPMFVSP 779
Qy 780 CVDKSKNPDFKVFYRSTLEKHLFISGLPFSCTKEELEECAGTGVKDLRLVNTNRAGK 839
Db 780 CVDKSKNPDFKVFYRSTLEKHLFISGLPFSCTKEELEECAGTGVKDLRLVNTNRAGK 839
Qy 840 PKGLAYVEYENESQASQAVKMDGMTIKENIKVAISNPPQKVPKPEKTRKAPGPMML 899
Db 840 PKGLAYVEYENESQASQAVKMDGMTIRENVIKVAISNPPQKVPKPEKTRKAPGPMML 899
Qy 900 POTYGARCKGTQSLPRALQPSAAPQAPGPAAPAAAPAAAPAAAPAAAPAAAPAAAPAA 959
Db 900 RMYGARGKGTQSLPRALQPSAAPQAPGPAAPAAAPAAAPAAAPAAAPAAAPAAAPAA 958
Qy 960 FLRK 963
Db 959 LLRK 962

RESULT 3
Q96J95
ID Q96J95 PRELIMINARY; PRT; 364 AA.
AC Q96J95;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DE KIA0156 Isoform.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Liu Y., He J.J.;
RT "Cloning and characterization of a novel Tat-binding protein.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF387506; AAK69347.1;
SQ SEQUENCE 364 AA; 41828 MW; D86B2562A7FA2E4E CRC64;

Query Match 36.3%; Score 1812; DB 4; Length 364;
Best Local Similarity 100.0%; Pred. No. 5.2e-101;
Matches 350; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MATAETSASPEAESKAGPADGEDEVKAARTRRKVLRSRAVAATYKTMGPWDQOE 60
Db 1 MATAETSASPEAESKAGPADGEDEVKAARTRRKVLRSRAVAATYKTMGPWDQOE 60
Qy 61 GVSESDGDEYAMASSAESSPGYEWYDEEENKQLETERLEEQSLINVDYNCHVDLIR 120
Db 61 GVSESDGDEYAMASSAESSPGYEWYDEEENKQLETERLEEQSLINVDYNCHVDLIR 120
Qy 121 LLRLEGELTKVRMAROKMSEIFPLTEELWLEWLHDEISMAQDGLDREHVDYDLFEKAVDY 180
Db 121 LLRLEGELTKVRMAROKMSEIFPLTEELWLEWLHDEISMAQDGLDREHVDYDLFEKAVDY 180
Qy 181 ICPNIWLEYGOYSGVGIGQKGLKVRVSVFRAALSSVGLHMTKGLALWEAYREFESAIVE 240
Db 181 ICPNIWLEYGOYSGVGIGQKGLKVRVSVFRAALSSVGLHMTKGLALWEAYREFESAIVE 240
```

```
Qy 241 AARLEKHSVLFRRQLAIPLYDMEATFAEYEWSEDPESVYQNTYKALQLEKYPYEE 300
Db 241 AARLEKHSVLFRRQLAIPLYDMEATFAEYEWSEDPESVYQNTYKALQLEKYPYEE 300
Qy 301 ALLOAEAPRLAYQAYIDFEMKIGDPARIQILFIERALVENCILVDPDLWIRY 350
Db 301 ALLOAEAPRLAYQAYIDFEMKIGDPARIQILFIERALVENCILVDPDLWIRY 350

RESULT 4
Q8S9I9
ID Q8S9I9 PRELIMINARY; PRT; 816 AA.
AC Q8S9I9;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DE ATG424270/T2A6_100.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Shinn P., Chen H., Cheuk R., Kim C.J., Meyers M.C., Banh J.,
RA Bowser L., Carninci P., Chang E., Dale J.M., Goldsmith A.D.,
RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,
RA Seki M., Southwick A., Tang C.C., Toriumi M., Wu H.C., Yamada K.,
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RT "Arabidopsis cDNA clones.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY075670; AAL77677.1;
SQ SEQUENCE 816 AA; 92808 MW; 2260EDC78469DBAF CRC64;

Query Match 18.1%; Score 902; DB 10; Length 816;
Best Local Similarity 28.3%; Pred. No. 6.5e-46;
Matches 271; Conservative 154; Mismatches 328; Indels 206; Gaps 34;

Qy 56 DQOEGVSESDGDEYAMASSAESSPGYEWYDEEENKQLETERLEEQSLINVDYNCH 115
Db 13 DQKMEASAPARADPPSDSDS--GSDSDSEDAESNH-QIVTLESELSANPYNDAY 69
Qy 116 VDLRLLEGELTKVRMAROKMSEIFPLTEELWLEWLHDEISMAQDGLDREHVDY--L 172
Db 70 VOYIKLLRTANLEKLRQAREMSAIFPLSPSLWLEWARDEASLAAS---ENVPEIVML 125
Qy 173 FEKAVKDYICPNWLEYGOY-----SVGGIGQKGLKVRVSVFRAALSSVGLHMTKGLA 226
Db 126 YERGLSDYQSVSLWCDYLSFMLEFDPSVRGYPSE-GISKMSLFEAIPAAGFHVTEG 184
Qy 227 LWAEYREFESAIVE-----EA-----ARLEKHSVLFRRQLAIPLYDMEATFAEYEW 272
Db 185 IWEGYREFEQGVLATIDEADIEERNKQIRINSIFRHLSPLENLSLSTLIAYKTWELEQ 244
Qy 273 -----SED--PIPESVYQNTYKALQLEKYPYEEALQ---AEAPRLAYQAYIDFE 320
Db 245 GIDLDIGSDLSKSHVSHQAVANKQAQMYSERAHLEENISKODLSDTSEKFEQFMYIKFE 304
Qy 321 MKIGDPARIQILFIERALVENCILVDPDLWIRYSOYLDRLQKLVKDLVLSINRNPWTVA 380
Db 305 KTSQDPTRVQAIYERAVAEYVPSSDLWIDYTVLDKTLKVGKAIATHAYSRATRSCPTGD 364
Qy 381 LWSRYLLAMERHGVHQVISTVFEKALNAGFTQATDYVEIMQAYLDYLRVRVDFKODSSK 440
Db 365 LWARYLLALERSASEKEIYDVFEKSLQCTSESSFEYLDLTRYDGLRRR----- 416
Qy 441 EELERAAFTRALEYLKVQVEERFNEGDPSCVIMON-----WARTEARLCNNMQ 490
Db 417 ----LSTRMLEALDY--SLIRETFQOASDYLTPHMONDTSLHLHTYTWANLELNGKDLA 470
```





[illegible]

```
Db 193 EIVFKNKIYDFHIDRLKLPHEOIEFTTSLSTFTVNNWSPSEY-EDVMYKSNKVYETTLK 251
QY 295 --YKPEEALLOAEAPRLAEYQAYIDFEMKIGDPA-----IQLIFERALVENCPLP 344
Db 252 RRAKIFNKELLNSANHSLE--AYMDL-----INDESRRTAEQLQYITTLERAILVPLIP 306
QY 345 DLWIRYSOYDLQRLKVLVSHNRAIRNCPTWVALWSRYLLAMERHGVHQVIVSTPE 404
Db 307 ELWLQYTAWLKSVDFSSQASSVAERATRNCSWIGRIWIKLTYMTLSGASTISAVCEEKD 366
QY 405 KALNAGFIQATDYVETWQAYLDYLRVRVDFKODSSKELEELRAAFTRALEYLKQVEERF 464
Db 367 RCLNSNLL--VNFDEVIDFSGFLKACLYLSSNEKDPQEFLLKHQIHKVEDYLK----- 418
QY 465 NESGDPSCVMQNAWRIEARL-----CANNOKARELWDSITMTGRNNAKYNMWLE--- 513
Db 419 NKGSKD-----AMRIELSKIYLYSISFESVEKWC--SDMFHDFQNALYISRYI 470
QY 514 ---YINLERAGHTQCRKALHRAVOCTSDYPHVCEVLLTMTERTSGSLEDWDIAVQKTE 570
Db 471 STMKYNPELA---AETLKSLYKNV---DQPQ-----LLQFYQSGINDLNDCFTNS 517
QY 571 TRLARVNEORM---RAAEKEALVQOEKEAKORRARAEEKALKKKKIRGPEKGADE 627
Db 518 HLYDVLNAORISFKRQDLSFAETKQTVENTPLKVPQADDTAALSKKRKPGQE----- 571
QY 628 DDEKEKGD--DEEOPKSKRRR-----VENSIPAAGETQNVFAAGPAGKCAAVDVEPPS 679
Db 572 -----GDVFKSKPIEQHNRREELTVLNTPLSDISENELKIFKDCGNIRIFILEDN 625
QY 680 KOKEKAASLK-----RDMPKVL-HDSSKD---SITVFVSNLPYSMQEPTDKLR 723
Db 626 QKDVKVAQIEFSETSEVLAAKTRDLKSIKRGHSIQIHVDVTNLYNTVFPPTVDELD--IT 683
QY 724 PLFEACGEVVOIRPFSNRGDFRGYCYVEFEKEKSALQALEMDRKSVEGR---PWFVSPC 780
Db 684 KLFSAYGNVVDVRRFSLRYNTNRRFCYVQMRKPDDEAHNALQJHKLLEEKYPIQVFIS-- 741
QY 781 VDKSNKPDPKFRYSTSLKHLFTISGLPFCSTKKELEICKAHGTVDKDLRLVTNRAGKP 840
Db 742 -----DPLRTRPSGAVYEGRELYNTIDFKVNEKDVETFFRDYGOVESVR-IPKRFNH 795
QY 841 KGLAVYENEYSQASQAVMKMDGMTIKENIKVAISNP 878
Db 796 KGFYGVVMTTQDAENA-LSAAGKQLGNRVNLNWLSPK 832
[1]
RESULT 8
Q9W4D2 PRELIMINARY; PRT; 941 AA.
AC Q9W4D2;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE RNPAF protein (L32008P).
GN RNPAF OR CG3312.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananatiades P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
```

```
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Gaig N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert X., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kianos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou S., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2193(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL Submitted (Oct-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF003433; AAF46022.1;
DR EMBL: AY061395; AAL28943.1;
DR HSSP: P09012; IURN.
DR FlyBase: FBgn0014024; Rnp4F.
DR InterPro: IPR003107; RNA.
DR InterPro: IPR000504; RNA_rec_mot.
DR Pfam: PF00076; rtm; 1.
DR SMART: SM00386; HAT; 3.
DR SMART: SM00360; RRM; 1.
DR PROSITE: PS00102; RRM; 1.
DR PROSITE: PS00030; RRM_RNP_1; 1.
SQ SEQUENCE 941 AA; 106739 MW; EAF2775B499FECFD CRC64;
Query Match 8.5%; Score 425; DB 5; Length 941;
Best Local Similarity 20.3%; Pred. No. 3.4e-17;
Matches 221; Conservative 151; Mismatches 347; Indels 372; Gaps 40;
QY 7 TSASPEAESKAGPKADGEDEVKAAARTRRKYLRSRAVAATYKTMGPAMDQOEQGESD 66
Db 87 SSDDEPSVEETEGGNAAGR-----GRANDSSSSDDVGVIE--GSELESNSESDDSD 138
QY 67 GDEYAMASSAESPGEYEWYDEEBEKNQLETERLEEQLSINVDYDNCVHDLRLRLLEG 126
Db 139 SD-----SDNAGGNGQLERSY---OELNALPSKKFAQWVS-----LIGIAFKLN 179
QY 127 ELTKVMARQKMSIEFFLTELWLELWLEHDEISMAQGLDREHVDLFEKAVKDYICPNW 186
Db 180 DLEKTESSVLEQLQNLATVPALVHLVKARLVVYVTTQDEERKAFPEOCARLEK 235
QY 187 LEYGOYSGVGIGQKGLKRVSRVFRALSSVGLHMTKGLALWEAVREFESAARLEK 246
Db 236 IPLSEYVNVYLVDQGNVQN-----HV-----LW-----AKLLA 263
```







21

```

DR InterPro: IPR000504; RNA_rec_mot.
DR Pfam: PF00076; rrm; 4.
DR SMART: SM00360; rrm; 4.
DR PROSITE: PS0102; RRM; 4.
DR PROSITE: PS00030; RRM_RNP_1; UNKNOWN_3.
SQ SEQUENCE 715 AA; 77439 MW; 5A420FAD8BAC6987 CRC64;

Query Match          5.1%; Score 252.5; DB 11; Length 715;
Best Local Similarity 24.9%; Pred. No. 5.2e-07;
Matches 101; Conservative 71; Mismatches 141; Indels 93; Gaps 16;

QY 553 ERTEGSLEDWDIAVQKTETRLARVNEQRKAA---EKEAALVQOEKEEAPQRKRAEKK 609
   . | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 198 EEDEDEDED-DEEEAMETTPAGKKAPVKAPVPAKAKSTAEDDEDEDEDEDEDEDEEE 256
   . | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 610 ALAKKKKIRPERKGADEDEKDWGDDDEEOP-----SKRRRVEN 649
   . : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 257 -----DEDDDEEEDDEEEEEEEPEYKEAPGKRKKEMAKQAKAPEAKKQVEA 305
   . : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 650 SIP-----AAGETQNVESAAGPAGKACAADVPEPPSQ 681
   . | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 306 TERTPTSFNLVGNLFSKSAPELKTGISDLFANKDLAVDVRIGVSRKKGYYDFE-SAED 364
   . KKAASL-----KRDMPKVLHDSSK--DSITVFNLSNLPYSMOEPTKLRPLFEAC 729
   . | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Db 365 LEKALEUTGLKVFNGNEIKLEPKPG-KDSKDRDARTLLAKNLPYKVTQDE--LKEVFE-- 419

QY 730 GEVVQIRPIFSNRGDFRGYCVFEKKEKSAQALEMDR-KSVGGRPMFVSPCYDVKSNPD 788
   . | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 420 -DAVEIR-LYSKDGKSGKIAYIEFKTEADAETFEKQGTGEIDGRSISLYTTGKGQSOD 477
   . | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 789 FKVFYRST-SLEKHKLFIISGLPFSCTKEELEETCKAHGTWKDLRLVTNRAGKPKGLAYVE 847
   . : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 478 YRGKKNSTWSGSEKTLVLSNLSYSATEETLQEVFEKATFIK---VPQNQNGSKGVAFIE 534
   . : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 848 YENESQASQAVMKMDGMTIKENILKVAISNPPORKVPEKPTRKAP 893
   . : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Db 535 FASFEQAKALNSCNREIGRAIRLFILOGP-----RGSPNARSQP 575
   . : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Search completed: June 18, 2003, 13:51:44  
Job time : 53 secs

Q9QZX1					
ID	Q9QZX1	PRELIMINARY;	PRT:	715 AA.	
AC	Q9QZX1;				
DT	01-MAY-2000	(TrEMBLrel. 13,	Created)		
DT	01-MAY-2000	(TrEMBurel. 13,	Last sequence update)		
DT	01-JUN-2002	(TrEMBurel. 21,	Last annotation update)		
DE	Nucleolin-related protein NRP.				
GN	NRP.				
OS	Rattus norvegicus (Rat).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.				
OX	NCBI_TaxID=10116;				
[1]	SEQUENCE FROM N.A.				
RN	[1]				
RP	STRAIN=SPRAGUE-DAWLEY;				
RC	MEDLINE=99419025; Pubmed=10488083;				
RX	Sorokina E.A., Kleinman J.G.;				
RA	"Cloning and preliminary characterization of a calcium-binding protein				
RT	closely related to nucleolin on the apical surface of inner medullary				
RT	collecting duct cells";				
RL	J. Biol. Chem. 274:27491-27496(1999).				
RL	EMBL; AF151373; AAD56625.1; -.				
DR	HSSP; P09651; 1HA1.				



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2003, 06:20:31 ; Search time 530 seconds  
(without alignments)  
16137.905 Million cell updates/sec

Title: US-09-763-985A-1

Perfect score: 3798

Sequence: 1 ccacgctcgcgtgacgact.....tcaaaaaaaaaaaaaaaaaa 3798

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_101002.\*  
1: /SID22/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SID22/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SID22/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SID22/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SID22/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SID22/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SID22/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SID22/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SID22/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SID22/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SID22/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SID22/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SID22/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SID22/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SID22/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SID22/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SID22/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SID22/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SID22/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SID22/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SID22/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SID22/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query	Match	Length	ID	Description
1	3798	100.0	3798	21	AAZ90783	Tumour antigen pro
2	2886	76.0	3537	22	AAD21574	Human SART3-IC-HER
3	2563.2	67.5	3355	21	AAF16121	Human prostate can
4	1331.2	35.1	1403	22	AAI98071	Human neuroblastom
5	1056.4	27.8	32145	22	AAI04631	Human reproductive
6	1056.4	27.8	32145	23	ABL97544	Human testicular a
7	436	11.5	1344	23	AAS91136	DNA encoding novel
8	398.2	10.5	437	21	AAA43952	Human secreted exp
9	393.4	10.4	432	24	ABL78750	Human ovarian can

10	392.4	10.3	430	24	ABL79050	Human ovarian can
11	346.8	9.1	554	22	AAH98082	Murine 7-transmemb
12	314.4	8.3	543	22	AAK75790	Human immune/haema
13	310.4	8.2	312	22	ABA49131	Human breast cell
14	310.4	8.2	312	22	ABA67045	Human foetal liver
15	310.4	8.2	312	22	ABA34138	Probe #12604 for g
16	310.4	8.2	312	22	ABA41586	Human brain expres
17	310.4	8.2	312	22	AAK1215	Human bone marrow
18	310.4	8.2	312	22	AAI21971	Probe #11904 for g
19	310.4	8.2	312	22	AAI47256	Probe #15942 used t
20	310.4	8.2	312	22	AAI07663	Probe #7654 used t
21	310.4	8.2	312	24	ABS15226	Human genome-deriv
22	291.2	7.7	320	21	AAC01859	Human secreted pro
23	233.4	6.1	324	24	ABN77683	Human ORF2630 cDNA
24	219.8	5.8	475	22	ABA44036	Human breast cell
25	219.8	5.8	475	22	ABA54492	Human foetal liver
26	219.8	5.8	475	22	ABA42476	Probe #2742 for ge
27	219.8	5.8	475	22	AAK02781	Human brain expres
28	219.8	5.8	475	22	AAK28223	Human bone marrow
29	219.8	5.8	475	22	AAI12792	Probe #2725 for ge
30	219.8	5.8	475	22	AAI34142	Probe #2828 used t
31	219.8	5.8	475	22	AAI02705	Probe #2696 used t
32	219.8	5.8	475	24	ABS02736	Human genome-deriv
33	170	4.5	170	14	AAQ61082	Human brain Expres
34	136.4	3.6	381	20	AAV90234	EST clone DF950.
35	129.6	3.4	185	22	ABA47794	Human breast cell
36	129.6	3.4	185	22	ABA65686	Human foetal liver
37	129.6	3.4	185	22	ABA32772	Probe #11238 for g
38	129.6	3.4	185	22	AAK14089	Human brain expres
39	129.6	3.4	185	22	AAK39833	Human bone marrow
40	129.6	3.4	185	22	AAI20643	Probe #10576 for g
41	129.6	3.4	185	22	AAI45853	Probe #14539 used
42	129.6	3.4	185	22	AAI06342	Probe #6333 used t
43	129.6	3.4	185	24	ABS13928	Human genome-deriv
44	126.4	3.3	467	22	ABA42665	Human breast cell
45	126.4	3.3	467	22	ABA53092	Human foetal liver

## ALIGNMENTS

### RESULT 1

AAZ90783  
ID AAZ90783 standard; DNA; 3798 BP.

XX AAZ90783;

XX AC AAZ90783;

XX DT 23-JUN-2000 (first entry)

XX DE Tumour antigen protein SART-3 encoding DNA.

XX DE Tumour antigen protein; SART-3; cytotoxic T-cell; HLA antigen;

XX KW tumour; ds.

XX OS Homo sapiens.

XX PN WO200012701-A1.

XX PD 09-MAR-2000.

XX PF 27-AUG-1999; 99WO-JP04622.

XX PR 28-AUG-1998; 98JP-0242660.

XX PA (SUMU ) SUMITOMO PHARM CO LTD.

XX PA (ITOHI/) ITOH K.

XX PI Itoh K, Nakao M;

XX DR WPI; 2000-237868/20.

XX DR P-PSDB; AAY85422.

XX PT Tumour antigen protein SART-3 recognized by cytotoxic T-cells binding to

PT HLA antigen for treatment and diagnosis of tumors

PS Claim 2; Page 58-63; 89pp; Japanese.

XX This DNA encodes a tumour antigen protein (SART-3) which is recognized by  
CC cytotoxic T-cells binding to HLA antigen. SART-3 can be expressed by  
CC standard recombinant methodology. SART-3 and its peptide derivatives can  
CC be used in the prevention, treatment and diagnosis of tumours in vitro or  
CC in vivo.

XX SQ Sequence 3798 BP; 1036 A; 847 C; 1074 G; 841 T; 0 other;

Query Match 100.0%; Score 3798; DB 21; Length 3798;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 3798; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	CCACGGCTCGGATGGCGCTGCGCGGAAACCTTCGGCTTCAGAACCCGAGGCTGAGTCCA	60
DB	1		
QY	61	AGGCTGGGCCCAAGGCTGACGGAGAGGAGGATGAGCTTAAGGCGGCTAGGACAAGGAGAA	120
DB	61		
QY	121	AGGTGTTATCGCGGGCTGTGGCGCTGCGACATACAAAGACCATGCGGGCCAGCGTGGGATC	180
DB	121		
QY	181	AGGTGTTATCGCGGGCTGTGGCGCTGCGACATACAAAGACCATGCGGGCCAGCGTGGGATC	180
DB	181		
QY	241	AGAGCTCCCGGGGAGTACGAGTGGGAATATGACGAAGAGGAGGAGAGAAAACCAAGCTGG	300
DB	241		
QY	301	AGATTGAGAGACTGGAGGAGAGTGTCTATCAACGCTCTATGACTACAACTGCCATGTGG	360
DB	301		
QY	361	ACTTGATCAGACTGCTCAGGCTGGAAGGGAGGCTTACCAAGTGAGGATGGCCGCCCA	420
DB	361		
QY	421	AGATGAGTGAATCTTTCCCTTGACTGAAGAGCTCTGGCTGGAGTGGCTGCATGACGAGA	480
DB	421		
QY	481	TCAGCATGCCAGGATGGCTGGACAGAGACGCTGTATGACCTCTTTTGAGAAAGCCG	540
DB	481		
QY	541	TGAAGATTACATTTGCTCTAATTTGGCTAGAGTATGCCAGTACTCAGTTGGTGGGA	600
DB	541		
QY	601	TTGGTCAGAAAGTGGCTTGAGAAAGTTGCGTCCGCTGTTGAAAGGGCTCTCTCGTCTG	660
DB	601		
QY	661	TTGGTTTACATATGACCAAGGACTCGCCCTCTGGAGGCTTACCGAGAGTTTGAAGTG	720
DB	661		
QY	721	CGATTCTGGAAGTCTGCTGGCTTGAGAAAGTCCACAGTCTTTTCGGGGCACAGTTGGCGA	780
DB	721		
QY	781	TCCCACTCTATGATATGGAGGCCACATTTGACAGATATGAAGATGGTTCAGAAAGCCCAA	840
DB	781		
QY	841	TACCAGAGTCAGTAATTCAGAACTATACAAAGCACTACAGCAGCTGGAGAAATATTAAC	900
DB	841		

DB	841	TACCAGAGTCAGTAATTCAGAACTATACAAAGCACTACAGCAGCTGGAGAAATATAAAC	900
QY	901	CCATATCAAGAAGCACTGTTGCAGGCAGAGCGCAAGCGTGCAGATATCAAGCATATA	960
DB	901		
QY	961	CTATATGAAGAAGCACTGTTGCAGGCAGAGCGCAAGCGTGCAGATATCAAGCATATA	960
DB	961		
QY	1021	TCGATTTTCAGATGAAATTTGGCGATCCCTGCTCGCATTCAGTTGATCTTTAGCGCGCC	1020
DB	1021		
QY	1081	TCGATTTTCAGATGAAATTTGGCGATCCCTGCTCGCATTCAGTTGATCTTTAGCGCGCC	1020
DB	1081		
QY	1141	TCGATTTTCAGATGAAATTTGGCGATCCCTGCTCGCATTCAGTTGATCTTTAGCGCGCC	1020
DB	1141		
QY	1201	TCGATTTTCAGATGAAATTTGGCGATCCCTGCTCGCATTCAGTTGATCTTTAGCGCGCC	1020
DB	1201		
QY	1261	TCGATTTTCAGATGAAATTTGGCGATCCCTGCTCGCATTCAGTTGATCTTTAGCGCGCC	1020
DB	1261		
QY	1321	TCGATTTTCAGATGAAATTTGGCGATCCCTGCTCGCATTCAGTTGATCTTTAGCGCGCC	1020
DB	1321		
QY	1381	TCGATTTTCAGATGAAATTTGGCGATCCCTGCTCGCATTCAGTTGATCTTTAGCGCGCC	1020
DB	1381		
QY	1441	TCGATTTTCAGATGAAATTTGGCGATCCCTGCTCGCATTCAGTTGATCTTTAGCGCGCC	1020
DB	1441		
QY	1501	TCGATTTTCAGATGAAATTTGGCGATCCCTGCTCGCATTCAGTTGATCTTTAGCGCGCC	1020
DB	1501		
QY	1561	TCGATTTTCAGATGAAATTTGGCGATCCCTGCTCGCATTCAGTTGATCTTTAGCGCGCC	1020
DB	1561		
QY	1621	TCGATTTTCAGATGAAATTTGGCGATCCCTGCTCGCATTCAGTTGATCTTTAGCGCGCC	1020
DB	1621		
QY	1681	TCGATTTTCAGATGAAATTTGGCGATCCCTGCTCGCATTCAGTTGATCTTTAGCGCGCC	1020
DB	1681		
QY	1741	TCGATTTTCAGATGAAATTTGGCGATCCCTGCTCGCATTCAGTTGATCTTTAGCGCGCC	1020
DB	1741		
QY	1801	TCGATTTTCAGATGAAATTTGGCGATCCCTGCTCGCATTCAGTTGATCTTTAGCGCGCC	1020
DB	1801		
QY	1861	TCGATTTTCAGATGAAATTTGGCGATCCCTGCTCGCATTCAGTTGATCTTTAGCGCGCC	1020
DB	1861		
QY	1921	TCGATTTTCAGATGAAATTTGGCGATCCCTGCTCGCATTCAGTTGATCTTTAGCGCGCC	1020
DB	1921		





QY 1512 ACCAGAGAAATGCCAAGTACGCCAACATGTGGCTAGAGTATTACAACTGGAAGAGCT 1571  
Db 1501 ACCAGAGAAATGCCAAGTACGCCAACATGTGGCTAGAGTATTACAACTGGAAGAGCT 1560  
QY 1572 CATGGTACACCCAGCACTGCCGAAGGCTCTGCACCGGGCCGTCCAGTGACACAGTGAC 1631  
Db 1561 CATGGTACACCCAGCACTGCCGAAGGCTCTGCACCGGGCCGTCCAGTGACACAGTGAC 1620  
QY 1632 TACCACAGACGCTCTGCGAAGTGTACTACCAATGGAGAGACAGAAGTGTCTTTAGAA 1691  
Db 1621 TACCACAGACGCTCTGCGAAGTGTACTACCAATGGAGAGACAGAAGTGTCTTTAGAA 1680  
QY 1692 GATTGGGATATAGCTGTTTCAAAAACCTGAACCCGATTTAGCTGTGTCATATGACAGAGA 1751  
Db 1681 GATTGGGATATAGCTGTTTCAAAAACCTGAACCCGATTTAGCTGTGTCATATGACAGAGA 1740  
QY 1752 ATGAAGGCTGCAGAGAAGGAGCAGCCCTTTGTGCAGCAAGAAAGAAAGGCTGAACAA 1811  
Db 1741 ATGAAGGCTGCAGAGAAGGAGCAGCCCTTTGTGCAGCAAGAAAGAAAGGCTGAACAA 1800  
QY 1812 CGGAAAGAGCTCGGGCTGAGAAGAAAGCGTTTAAAAAGAAAGAAAGATCAGAGGCCCA 1871  
Db 1801 CGGAAAGAGCTCGGGCTGAGAAGAAAGCGTTTAAAAAGAAAGAAAGATCAGAGGCCCA 1860  
QY 1872 GAGAAAGCGGAGCAGATGAGGACGATGAGAAAGATGGGGCGATGATGAAGAAGAGCAG 1931  
Db 1861 GAGAAAGCGGAGCAGATGAGGACGATGAGAAAGATGGGGCGATGATGAAGAAGAGCAG 1920  
QY 1932 CCTTCCAAACCCAGAAAGGTCGAGNACAGCATCCCTGCAGCTGGAGAAACACAAATGTA 1991  
Db 1921 CCTTCCAAACCCAGAAAGGTCGAGNACAGCATCCCTGCAGCTGGAGAAACACAAATGTA 1980  
QY 1992 GAAGTAGCAGCAGGGCCCGCTGGAAATGTCTGCCGTAGATGTGGAGCCCTTCGGAAG 2051  
Db 1981 GAAGTAGCAGCAGGGCCCGCTGGAAATGTCTGCCGTAGATGTGGAGCCCTTCGGAAG 2040  
QY 2052 CAGAAGGAGAGGAGCAGCTCCCTGAAGAGGACATGCCAAAGGTGCTGCAGCAGACGAGC 2111  
Db 2041 CAGAAGGAGAGGAGCAGCTCCCTGAAGAGGACATGCCAAAGGTGCTGCAGCAGACGAGC 2100  
QY 2112 AAGNACAGCATCAGCGTCTTTGTCTAGCAACCTGCCCTACAGATCGAGGAGCGGACAG 2171  
Db 2101 AAGNACAGCATCAGCGTCTTTGTCTAGCAACCTGCCCTACAGATCGAGGAGCGGACAG 2160  
QY 2172 AAGCTCAGGCCACTCTTCAGGCGCTGTGGGAGGTGTCCAGATCCGACCCATCTTCAGC 2231  
Db 2161 AAGCTCAGGCCACTCTTCAGGCGCTGTGGGAGGTGTCCAGATCCGACCCATCTTCAGC 2220  
QY 2232 AACCGTGGGATTTCCGAGGTTACTGCTACGTGGAGTTTAAAGAGAGAAATCAGCCCTT 2291  
Db 2221 AACCGTGGGATTTCCGAGGTTACTGCTACGTGGAGTTTAAAGAGAGAAATCAGCCCTT 2280  
QY 2292 CAGGCACCTGGAGATGGACCGGAAAGTGTAGAAGGAGCGCAATGTTTGTCCCCCTGT 2351  
Db 2281 CAGGCACCTGGAGATGGACCGGAAAGTGTAGAAGGAGCGCAATGTTTGTCCCCCTGT 2340  
QY 2352 GTGGATAAGAGCAAAACCCGATTTTAAAGTGTTCAGGTACAGCACTTCCTCTAGAGAA 2411  
Db 2341 GTGGATAAGAGCAAAACCCGATTTTAAAGTGTTCAGGTACAGCACTTCCTCTAGAGAA 2400  
QY 2412 CACAAGCTGTTTCACTTCAGGCGCTGCCCTTCTCTACTTAAAGAGAACTAGAAGAAATC 2471  
Db 2401 CACAAGCTGTTTCACTTCAGGCGCTGCCCTTCTCTACTTAAAGAGAACTAGAAGAAATC 2460  
QY 2472 TGTAAGGCTCATGCGACCGTGAAGACCTCAGGCTGTCCACCAACCGGGCTGGCAACCA 2531  
Db 2461 TGTAAGGCTCATGCGACCGTGAAGACCTCAGGCTGTCCACCAACCGGGCTGGCAACCA 2520  
QY 2532 AAGGCGCTGGCTACGTGAGGTATGAAATGAATCCAGGGGTCCAGGCTGTGTATGAAG 2591  
Db 2521 AAGGCGCTGGCTACGTGAGGTATGAAATGAATCCAGGGGTCCAGGCTGTGTATGAAG 2580  
QY 2592 ATGGACGGCATGACTATCAAGAGAAACATCATCAAAAGTGGCAATCAGCAACCCCTCCTCAG 2651

Db 2581 ATGGACGGCATGACTATCAAGAGAAACATCATCAAAAGTGGCAATCAGCAACCCCTCCTCAG 2640  
QY 2652 AGGAAAGTTCAGAGAACCCAGACAGACAGGAGGACCAAGTGGCCCATGCTTTTGGCG 2711  
Db 2641 AGGAAAGTTCAGAGAACCCAGACAGACAGGAGGACCAAGTGGCCCATGCTTTTGGCG 2700  
QY 2712 CAGACATACGGAGCGAGGGGGAAGGAGGAGCGAGCTGTCTTACTTGTCTGCTGCGCCCTG 2771  
Db 2701 CAGACATACGGAGCGAGGGGGAAGGAGGAGCGAGCTGTCTTACTGCTGCTGCGCCCTG 2760  
QY 2772 CAGCGCCCAAGTGTGTGAGCTCCTCAGGCTGAGAACGGCCCTGCGCGGCTCCTGAGTT 2831  
Db 2761 CAGCGCCCAAGTGTGTGAGCTCCTCAGGCTGAGAACGGCCCTGCGCGGCTCCTGAGTT 2820  
QY 2832 GCCGCCCCAGCAGCCACCGAGGACCCCAAGATGTCCAATGCCGATTTTGCCCAAGCTGTTT 2891  
Db 2821 GCCGCCCCAGCAGCCACCGAGGACCCCAAGATGTCCAATGCCGATTTTGCCCAAGCTGTTT 2880  
QY 2892 CTGAGA 2897  
Db 2881 CTGAGA 2886  
RESULT 3  
AAF16121  
ID AAF16121 standard; cDNA; 3355 BP.  
XX AAF16121;  
AC AAF16121;  
DT 13-MAR-2001 (first entry)  
XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:556.  
DE Human; prostate cancer; prostate cancer antigen; detection; diagnosis;  
XX neuroprotective; cytosolic; cardioactive; immunomodulatory; muscular;  
KW vulnary; gastrointestinal; nephrotropic; antiinfective; gynaecological;  
KW antibacterial; gene therapy; neural; immune; reproductive; renal;  
KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;  
KW wound; infectious disease; ss.  
XX Homo sapiens.  
OS  
XX WO200055174-A1.  
PN  
XX 21-SEP-2000.  
PD  
XX 08-MAR-2000; 2000MO-US05988.  
PF  
XX 12-MAR-1999; 99US-0124270.  
PR  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA (ROSE/) ROSEN C A.  
XX  
PI Rosen CA, Ruben SM;  
XX  
DR WPI: 2000-587513/55.  
DR P-PSDB; AAB56918.  
XX  
PT Prostate cancer associated gene sequences, referred to as prostate  
PT cancer antigens, useful for treatment, prevention, and diagnosis of  
PT disorders such as prostate cancer -  
XX  
XX Claim 1; Page 1014-1015; 2338pp; English.  
PS  
XX AAF15566 to AAF16505 encode the human prostate cancer associated  
CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.  
CC The prostate cancer antigens can have neuroprotective, cytostatic,  
CC cardioactive, immunomodulatory, muscular, vulnary, gastrointestinal,  
CC nephrotropic, antiinfective, gynaecological and antibacterial activities,  
CC and can be used in gene therapy. The prostate cancer antigen  
CC polynucleotides may be used for detection of prostate cancer, chromosome  
CC identification, as chromosome markers, and for numerous other diagnostic











AC ABL97544;  
XX 21-JUN-2002 (first entry)  
XX Human testicular antigen encoding DNA fragment SEQ ID NO: 2196.  
XX  
XX  
XX Human; testicular antigen; testes; cancer; metastasis; immune disorder;  
KW reproductive system disorder; urinary system disorder; gene therapy;  
KW cardiovascular disorder; respiratory disorder; neurological disorder;  
KW gastrointestinal disease; infection; cytostatic; gene; ds.  
XX  
XX Homo sapiens.  
XX WO200155317-A2.  
XX  
XX 02-AUG-2001.  
XX  
XX 17-JAN-2001; 2001WO-US01329.  
XX  
XX 31-JAN-2000; 2000US-0179065.  
XX 04-FEB-2000; 2000US-0180628.  
XX 24-FEB-2000; 2000US-0184664.  
XX 02-MAR-2000; 2000US-0186350.  
XX 15-MAR-2000; 2000US-0189874.  
XX 17-MAR-2000; 2000US-0190076.  
XX 18-APR-2000; 2000US-0198123.  
XX 19-MAY-2000; 2000US-0205515.  
XX 07-JUN-2000; 2000US-0209467.  
XX 28-JUN-2000; 2000US-0214886.  
XX 30-JUN-2000; 2000US-0215135.  
XX 07-JUL-2000; 2000US-0216647.  
XX 07-JUL-2000; 2000US-0216880.  
XX 11-JUL-2000; 2000US-0217487.  
XX 11-JUL-2000; 2000US-0217496.  
XX 14-JUL-2000; 2000US-0218290.  
XX 26-JUL-2000; 2000US-0220963.  
XX 26-JUL-2000; 2000US-0220964.  
XX 14-AUG-2000; 2000US-0224518.  
XX 14-AUG-2000; 2000US-0225213.  
XX 14-AUG-2000; 2000US-0225214.  
XX 14-AUG-2000; 2000US-0225266.  
XX 14-AUG-2000; 2000US-0225267.  
XX 14-AUG-2000; 2000US-0225268.  
XX 14-AUG-2000; 2000US-0225270.  
XX 14-AUG-2000; 2000US-0225447.  
XX 14-AUG-2000; 2000US-0225757.  
XX 14-AUG-2000; 2000US-0225758.  
XX 14-AUG-2000; 2000US-0225759.  
XX 18-AUG-2000; 2000US-0226279.  
XX 22-AUG-2000; 2000US-0226681.  
XX 22-AUG-2000; 2000US-0226868.  
XX 22-AUG-2000; 2000US-0227182.  
XX 23-AUG-2000; 2000US-0227009.  
XX 30-AUG-2000; 2000US-0228924.  
XX 01-SEP-2000; 2000US-0229287.  
XX 01-SEP-2000; 2000US-0229343.  
XX 01-SEP-2000; 2000US-0229344.  
XX 01-SEP-2000; 2000US-0229345.  
XX 05-SEP-2000; 2000US-0229509.  
XX 05-SEP-2000; 2000US-0229513.  
XX 06-SEP-2000; 2000US-0230437.  
XX 06-SEP-2000; 2000US-0230438.  
XX 08-SEP-2000; 2000US-0231242.  
XX 08-SEP-2000; 2000US-0231243.  
XX 08-SEP-2000; 2000US-0231244.  
XX 08-SEP-2000; 2000US-0231413.  
XX 08-SEP-2000; 2000US-0231414.  
XX 08-SEP-2000; 2000US-0232081.  
XX 08-SEP-2000; 2000US-0232081.  
XX 12-SEP-2000; 2000US-0231968.  
XX 14-SEP-2000; 2000US-0232397.  
XX 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 14-SEP-2000; 2000US-0233066.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 01-DEC-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.





QY 393 CTTACCAAGGTGAGTGGCCGCCAGAGATGAGTGAATCTTTCCCTTGACTGAAGAG 452  
D 141 CTTACCAAGGTGAGTGGCCGCCAGAGATGAGTGAATCTTTCCCTTGACTGAAGAG 200  
QY 453 CTCGGCTGGAGTGGCTGCATGACGAGATCAGCATGCGCCAGGATGGCCTGGACAGAG 512  
D 201 CTCGGCTGGAGTGGCTGCATGACGAGATCAGCATGCGCCAGGATGGCCTGGACAGAG 260  
QY 513 CACGTGTATGACCTCTTTGAGAAAGCGTGAAGATTACATTTCTCTAACATTTGGCTA 572  
D 261 CACGTGTATGACCTCTTTGAGAAAGCGTGAAGATTACATTTCTCTAACATTTGGCTA 320  
QY 573 GAGTATGCCCACTACTCAGTTGGTGGATTGGTGCAGAAAGTGCGCTTGAGAAAGTTCGC 632  
D 321 GAGTATGCCCACTACTCAGTTGGTGGATTGGTGCAGAAAGTGCGCTTGAGAAAGTTCGC 380  
QY 633 TCCGTGTTGAAAGGGCTCTCTCGTCTGTTGGTTTACATATGACCAAGGACTCG 687  
D 381 TCCGTGTTGAAAGGGCTCTCTCGTCTGTTGGTTTACATATGACCAAGGACTCG 435

## RESULT 9

ABL78750  
ID ABL78750 standard; cDNA; 432 BP.

XX ABL78750;

XX 17-MAY-2002 (first entry)

XX Human ovarian cancer related cDNA clone SEQ ID NO:1728.

XX Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.

XX Homo sapiens.

XX WO200192581-A2.

XX 06-DEC-2001.

XX 29-MAY-2001; 2001WO-US17756.

XX 26-MAY-2000; 2000US-207484P.

XX (CORI-) CORIXA CORP.

XX Algate PA, Harlocker SL, Jones R;

XX WPI; 2002-122075/16.

XX Composition for therapy and diagnosis of ovarian cancer comprising polypeptide of a ovarian tumor polypeptide, polynucleotide encoding polypeptide, antibody specific to polypeptide or T cell expressing polypeptide -

PS Claim 1; SEQ ID 1728; 489pp; English.

XX The present invention describes a composition (I) comprising: carriers and immunostimulants; and a polypeptide (II) of a ovarian tumour polypeptide encoded by a polynucleotide (III) having a cDNA sequence (S1) from the 10912 nucleotide sequences as given in ABL77023 to ABL87934, (III) encoding (II) having a sequence (S2), a T cell population of (II), or antigen presenting cells that express (II). (I) has cytostatic activity. An oligonucleotide (IV) that hybridises to (S1) can be used for detecting ovarian cancer in a patient's biological sample preferably serum or ovarian tissue. The method comprises contacting a biological sample from a patient with (IV), detecting the amount of polynucleotide hybridising to (IV) and comparing the amount to a predetermined cutoff value and thereby detecting ovarian cancer in the patient, where the amount of polynucleotide hybridising to (IV) is detected preferably by polymerase chain reaction (PCR). (I) comprising (III) and/or (II) is useful for stimulating and/or expanding T cells specific for an ovarian tumour protein comprising contacting T cells

CC with (III) or (II). (III) is useful in design and preparation of ribozyme molecules for inhibiting expression of the tumour polypeptides and proteins in tumour cells; and to isolate a full length gene from a suitable library e.g., a tumour cDNA library using well known techniques.

XX SQ Sequence 432 BP; 108 A; 86 C; 105 G; 132 T; 1 other;

Query Match 10.4%; Score 393.4; DB 24; Length 432;  
Best Local Similarity 98.8%; Pred. No. 2.1e-96;  
Matches 427; Conservative 0; Mismatches 2; Indels 3; Gaps 3;

QY 3063 CAAGAGCCTTTCAGTGTCTCCCTCATATTGAGGCGAGTGGCAGAAAGTGACCACCTCTGCA 3122  
D 2 CAAGAGCCTTTCAGTGTCTCCCTCATATTGAGGCGAGTGGCAGAAAGTGACCACCTCTGCA 61

QY 3123 GGCTGGGCCAGGATGGTGTCTCTGAGATAGTTTGTATCTTAAAGACTGAGGCACAGA 3182

D 62 GGCTGGG-CCAGGATGGTGTCTCTGAGATAGTTTGTATCTTAAAGACTGAGGCACAGA 120

QY 3183 AGCGAAACGAGAACACA-CTGTTTTTGAGACACAGTTCTCCAAATGTTTCTGGCCAGCTC 3241

D 121 AGCGAAACGAGAACACACCTGTTTTGAGACACAGTTCTCCAAATGTTTCTGGCCAGCTC 180

QY 3242 CGGCCCTTTTGTATGACACTTCTTCCACCTGCACAGCACATGTGCCGTCATTCT 3301

D 181 CGGCCCTTTTGTATGACACTTCTTCCACCTGCACAGCACATGTGCCGTCATTCT 240

QY 3302 TTTAATTTTAAAGATGAAATGGCAGATGCTAGTAATTCACAGAAATGCCCTCTTGTGGGG 3361

D 241 TTTAATTTTAAAGATGAAATGGCAGATGCTAGTAATTCACAGAAATGCCCTCTTGTGGGG 300

QY 3362 GTGGGCTGCGAGGAGTCAGCTATAAAACATTTCTGCGAGTTTCTCAATGGGCTCTG 3421

D 301 GTGGGCTGCGAGGAGTCAGCTATAAAACATTTCTGCGAGTTTCTCAATGGGCTCTG 360

QY 3422 CATTTT-ATATTATGTTTGTAAATGACATGTGAGCCCTTGTTCATGTTTCTCAAAA 3480

D 361 CATTTTAAATATTATGTTTGTAAATGACATGTGAGCCCTTGTTCATGTTTCTCAAAA 420

QY 3481 GCAGAAATTTG 3492

D 421 GCAGAAATTTG 432

## RESULT 10

ABL79050  
ID ABL79050 standard; cDNA; 430 BP.

XX ABL79050;

XX 17-MAY-2002 (first entry)

XX Human ovarian cancer related cDNA clone SEQ ID NO:2028.

XX Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.

XX Homo sapiens.

XX WO200192581-A2.

XX 06-DEC-2001.

XX 29-MAY-2001; 2001WO-US17756.

XX 26-MAY-2000; 2000US-207484P.

XX (CORI-) CORIXA CORP.

XX Algate PA, Harlocker SL, Jones R;

XX WPI; 2002-122075/16.



PT Composition for therapy and diagnosis of ovarian cancer comprising  
PT polypeptide of an ovarian tumor polypeptide, polynucleotide encoding  
PT polypeptide, antibody specific to polypeptide or T cell expressing  
XX polypeptide  
PS Claim 1; SEQ ID 2028; 489pp; English.  
XX The present invention describes a composition (I) comprising: carriers  
CC and immunostimulants; and a polypeptide (II) of an ovarian tumor  
CC polypeptide encoded by a polynucleotide (III) having a cDNA sequence  
CC (S1) from the 10912 nucleotide sequences as given in ABL77023 to  
CC ABL87934, (III) encoding (II) having a sequence (S2), a T cell  
CC population of (II), or antigen presenting cells that express (II).  
CC (I) has cytostatic activity. An oligonucleotide (IV) that hybridises to  
CC (S1) can be used for detecting ovarian cancer in a patient's biological  
CC sample preferably serum or ovarian tissue. The method comprises  
CC contacting a biological sample from a patient with (IV), detecting the  
CC amount of polynucleotide hybridising to (IV) and comparing the amount to  
CC a predetermined cutoff value and thereby detecting ovarian cancer in the  
CC patient, where the amount of polynucleotide hybridising to (IV) is  
CC detected preferably by polymerase chain reaction (PCR). (I) comprising  
CC (III) and/or (II) is useful for stimulating and/or expanding T cells  
CC specific for an ovarian tumor protein comprising contacting T cells  
CC with (III) or (II). (III) is useful in design and preparation of  
CC ribozyme molecules for inhibiting expression of the tumour polypeptides  
CC and proteins in tumour cells; and to isolate a full length gene from a  
CC suitable library e.g., a tumour cDNA library using well known  
CC techniques.  
XX  
SQ Sequence 430 BP; 108 A; 85 C; 104 G; 132 T; 1 other;

Query Match 10.3%; Score 392.4; DB 24; Length 430;  
Best Local Similarity 98.8%; Pred. No. 4e-96;  
Matches 426; Conservative 0; Mismatches 2; Indels 3; Gaps 3;  
QY 3064 AAGAGCCTTTCAGTCTCCTCATATGAGGCGAGTGGCAGAAAGTGACCACTCTGCAG 3123  
DB 1 AAGAGCCTTTCAGTCTCCTCATATGAGGCGAGTGGCAGAAAGTGACCACTCTGCAG 60  
QY 3124 GCTGGGCCAGGATGTGTCTCTGAGATAGTGTGTATCTTAAAGACTGAGGCACAGAA 3183  
DB 61 GCTGGG-CCAGATGTGTGTCTCTGAGATAGTGTGTATCTTAAAGACTGAGGCACAGAA 119  
QY 3184 GCGAAGCAGAGACACA-CGTGTTTTGAGACAGATGTGCAAAATGTTCTGCCAGCTCC 3242  
DB 120 GCGAAGCAGAGACACACCTGTTTTGAGACAGATGTGCAAAATGTTCTGCCAGCTCC 179  
QY 3243 GCGCCCTTTTCTATGACACTTCTCTCCACCTGCACAGACATGTGCCCGCTCATCTT 3302  
DB 180 GCGCCCTTTTCTATGACACTTCTCTCCACCTGCACAGACATGTGCCCGCTCATCTT 239  
QY 3303 TTAATTTTAAAGATGAATGCGAGATGCTAGTAATTCACAGAAATGGCCTCTGTGGGG 3362  
DB 240 TTAATTTTAAAGATGAATGCGAGATGCTAGTAATTCACAGAAATGGCCTCTGTGGGG 299  
QY 3363 TGGGTCTGAGGAAGTACAGCTATAAACAATTTGCTGGAGTTTTGTTCAATGGGCTGTGC 3422  
DB 300 TGGGTCTGAGGAAGTACAGCTATAAACAATTTGCTGGAGTTTTGTTCAATGGGCTGTGC 359  
QY 3423 ATTTT-ATATTATGTTTGTAAATGACATGTCAGCCCTGTTTCATGTTTCCCTAAAG 3481  
DB 360 ATTTTAAATATGTTTGTAAATGACATGTCAGCCATTTGTTTCATGTTTCCCTAAAG 419  
QY 3482 CAGAATATTG 3492  
DB 420 CAGAATATTG 430

RESULT 11  
ID AAH98082  
XX AAH98082 standard; DNA; 554 BP.  
AC AAH98082;

XX 10-OCT-2001 (first entry)  
XX Murine 7-transmembrane G-protein coupled receptor coding sequence #326.  
XX  
XX Murine; stromal stem cell; signalling; vaccine; 7TM-GPCR;  
KW 7-transmembrane G-protein coupled protein receptor; ds.  
XX  
OS Mus sp.  
XX WO200160999-A1.  
XX 23-AUG-2001.  
XX 14-FEB-2001; 2001WO-US04700.  
XX 14-FEB-2000; 2000US-0182377.  
XX (IMCL-) IMCLONE SYSTEMS INC.  
PA (UYPR-) UNIV PRINCETON.  
XX Lemischka IR, Witte L, Pereira DS;  
XX WPI; 2001-522596/57.  
XX DNA Sequences encoding 7-transmembrane G-protein coupled protein  
PT receptors characteristic of hematopoietic stem cells, useful for  
PT treating leukemia.  
XX  
PS Claim 1; Page 143-144; 176pp; English.  
XX The present invention relates to murine coding sequences for  
CC 7-transmembrane G-protein coupled protein receptors (7TM-GPCRs). The  
CC present sequence is one such murine 7TM-GPCR coding sequence. The present  
CC sequence was derived from stromal stem cells. The present sequence  
CC and its corresponding protein are useful in the prevention, diagnosis and  
CC treatment of diseases associated with inappropriate 7TM-GPCR expression.  
CC 7TM-GPCRs identify specific signalling molecules, to activate an  
CC effector-signalling cascade that triggers an intracellular response and  
CC eventually a biological effect.  
XX  
SQ Sequence 554 BP; 126 A; 127 C; 162 G; 124 T; 15 other;

Query Match 9.1%; Score 346.8; DB 22; Length 554;  
Best Local Similarity 79.9%; Pred. No. 1.2e-83;  
Matches 436; Conservative 0; Mismatches 106; Indels 4; Gaps 3;  
QY 559 CTAACATTTGGCTAGATATGCCAGTACTCAGTTGGTGGGATGGTCAGAAGGTGGCC 618  
DB 1 CAACATTTGGCTAGATATGCCAGTACTCAGTTGGTGGGATGGTCAGAAGGTGGCC 60  
QY 619 TTGAGAAAGTTCCGCTCCGCTGTTTGAAGGGCTCTCTCGTCTGTTGTTTACATATGACCA 678  
DB 61 TTGAGAAAGTTCCGCTCTGCTCTTTGAAGAGCCCTGCTCTGTTGGCTGCACATGACCA 120  
QY 679 AAGSACTCCGCTCTGGAGGCTTACCGAGAGTTTGAAGTCGATTTGGAAGTGTCTC 738  
DB 121 AAGGCTCGCCATCTGGGAGGCTTACCGANAGTTTGAAGCGCCATCGTGGAGGCTGCTC 180  
QY 739 GCGTTGAGAAAGTCCACAGTCTTTTCGCGGACAGTGGCGATCCCACTCTATGATAGG 798  
DB 181 GGCTGGAGAAAGTGACAGTCTCTTTCCGCGGACA-CTNCGCATCCCACTGTACGAGATGG 239  
QY 799 AGGCCACATTTTCAGAGTATGAAGATGGTCAGAAGCCCAATACAGATCAGTAATTC 858  
DB 240 AGGCCACCTTTTCAGAGTATGAAGATGGTCAGAGGCCCAATGCGGNGTCTGTACTTC 299  
QY 859 AGAAGTATAAAGCACTACAGAGCTGGAGAAATATAAACCCTATGAAGAGCACTGT 918  
DB 300 AGNGTATCAGAAAGCGCTGGGCGAGCTANAGAGTACAAAGCTTACGANGAAGCGCTGC 359  
QY 919 TCCAGGAGGCCACCAAGGCTGGCAGATATCAAGCATATATCATTTTGGATGAGAA 978  
|||||

```
Db 360 TGCAGCAGANGCCCTCGGNGCGGAATACCAAGCTTACATCGANTTCGAGATGAAA 419
Qy 979 TTGGCGATCCTGCTCGCATTCAGTTGATCTTTGAGCGGCCCTGGTCGAGACTGCGTTG 1038
Db 420 TCGGGGAT-CTGGCCGATTACGTTGATCTTTGAGCGTTCTGGTGGAGAACTTGCCTGG 478
Qy 1039 TCCAGACTTATGATCCGTTTACAGTCAGTACCTAGATCGACAACCTGAAAAGTAAGGATT 1098
Db 479 NTCCAGNCTTATGATC- GCCACAGTNGTACCTAGATCCACAGCTGAAGTTAAGGCT 536
Qy 1099 TGGTTT 1104
Db 537 TGGTTT 542

RESULT 12
AAK75790/c
ID AAK75790 standard; DNA; 543 BP.
XX AC
XX AC
XX AC
XX DT
XX 07-NOV-2001 (first entry)
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:30602.
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX Homo sapiens.
XX OS
XX WO200157182-A2.
XX PN
XX PD
XX 09-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US01354.
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
```



CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 312 BP; 41 A; 125 C; 66 G; 80 T; 0 other;

Query Match 8.2%; Score 310.4; DB 22; Length 312;  
Best Local Similarity 99.7%; Pred. No. 6.5e-74;  
Matches 311; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 12 ATGCGACTGCGGCCGGAACCTCGCTTCAGAACCCGAGGCTGAGTCCAAAGGCTGGGCC 71  
Db 312 ATGCGACTGCGGCCGGAACCTCGCTTCAGAACCCGAGGCTGAGTCCAAAGGCTGGGCC 253  
Qy 72 AAGCTGACGAGAGGAGGATGAGTTAAGCGGCTAGGACAAGGAGAAAGTTATCG 131  
Db 252 AAGCTGACGAGAGGAGGATGAGTTAAGCGGCTAGGACAAGGAGAAAGTTATCG 193  
Qy 132 CGGCTGTGGCCGCTGCGACATACAGACCATGGGCCAGCGTGGGATCAGCAGAGGAA 191  
Db 192 CGGCTGTGGCCGCTGCGACATACAGACCATGGGCCAGCGTGGGATCAGCAGAGGAA 133  
Qy 192 GGCCTGAGCAGAGCGATGGGATGAGTACGCCATGGCTTCCTCCGCGAGAGCTCCGCC 251  
Db 132 GGCCTGAGCAGAGCGATGGGATGAGTACGCCATGGCTTCCTCCGCGAGAGCTCCGCC 73  
Qy 252 GGGGAGTACGAGTGGGAATATGACGAAGAGAGAGAGAAACCAGCTGGAGATTGAGAGA 311  
Db 72 GGGGAGTACGAGTGGGAATATGACGAAGAGAGAGAGAAACCAGCTGGAGATTGAGAGA 13  
Qy 312 CTGGAGGAGCAG 323  
Db 12 CTGGAGGAGCAG 1

## RESULT 14

ABA67045/C  
ID ABA67045 standard; DNA; 312 BP.

XX  
AC ABA67045;

XX  
DT 01-FEB-2002 (first entry)

XX Human foetal liver single exon nucleic acid probe #15350.

XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.

XX Homo sapiens.

XX WO200157277-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00669.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483447/52.

XX Human genome-derived single exon nucleic acid probes useful for  
XX analyzing gene expression in human foetal liver.

XX Claim 4; SEQ ID NO 15350; 639pp + sequence listing; English.

XX The invention relates to a single exon nucleic acid probe for

CC measuring human gene expression in a sample derived from human foetal  
CC liver. The single exon nucleic acid probes may be used for predicting,  
CC measuring and displaying gene expression in samples derived from human  
CC foetal liver. The present sequence is a single exon nucleic acid  
CC probe of the invention.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 312 BP; 41 A; 125 C; 66 G; 80 T; 0 other;

Query Match 8.2%; Score 310.4; DB 22; Length 312;  
Best Local Similarity 99.7%; Pred. No. 6.5e-74;  
Matches 311; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 12 ATGCGACTGCGGCCGGAACCTCGCTTCAGAACCCGAGGCTGAGTCCAAAGGCTGGGCC 71  
Db 312 ATGCGACTGCGGCCGGAACCTCGCTTCAGAACCCGAGGCTGAGTCCAAAGGCTGGGCC 253  
Qy 72 AAGCTGACGAGAGGAGGATGAGTTAAGCGGCTAGGACAAGGAGAAAGTTATCG 131  
Db 252 AAGCTGACGAGAGGAGGATGAGTTAAGCGGCTAGGACAAGGAGAAAGTTATCG 193  
Qy 132 CGGCTGTGGCCGCTGCGACATACAGACCATGGGCCAGCGTGGGATCAGCAGAGGAA 191  
Db 192 CGGCTGTGGCCGCTGCGACATACAGACCATGGGCCAGCGTGGGATCAGCAGAGGAA 133  
Qy 192 GGCCTGAGCAGAGCGATGGGATGAGTACGCCATGGCTTCCTCCGCGAGAGCTCCGCC 251  
Db 132 GGCCTGAGCAGAGCGATGGGATGAGTACGCCATGGCTTCCTCCGCGAGAGCTCCGCC 73  
Qy 252 GGGGAGTACGAGTGGGAATATGACGAAGAGAGAGAGAAACCAGCTGGAGATTGAGAGA 311  
Db 72 GGGGAGTACGAGTGGGAATATGACGAAGAGAGAGAGAAACCAGCTGGAGATTGAGAGA 13  
Qy 312 CTGGAGGAGCAG 323  
Db 12 CTGGAGGAGCAG 1

## RESULT 15

ABA34138/C  
ID ABA34138 standard; DNA; 312 BP.

XX  
AC ABA34138;

XX  
DT 23-JAN-2002 (first entry)

XX Probe #12604 for gene expression analysis in human heart cell sample.

XX Human; gene expression; heart; microarray; vascular system; probe;  
XX cardiovascular disease; hypertension; cardiac arrhythmia;  
XX congenital heart disease; ss.

XX Homo sapiens.

XX WO200157277-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00666.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-48899/53.  
XX Single exon nucleic acid probes for analyzing gene expression in human  
PT hearts  
XX  
XX  
PS Claim 4; SEQ ID No 12604; 530pp; English.  
XX  
XX The present invention relates to single exon nucleic acid probes for  
CC measuring human gene expression in a sample derived from human heart. The  
CC present sequence is one such probe. The probes may be used for  
CC predicting, measuring and displaying gene expression in samples derived  
CC from the human heart via microarrays. By measuring gene expression, the  
CC probes are useful for predicting, diagnosing, grading, staging,  
CC monitoring and prognosing diseases of the human heart and vascular system  
CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and  
CC congenital heart disease.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 312 BP; 41 A; 125 C; 66 G; 80 T; 0 other;

Query Match 8.2%; Score 310.4; DB 22; Length 312;  
Best Local Similarity 99.7%; Pred. No. 6.5e-74;  
Matches 311; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 12 ATGGCGACTGGCGCGAACCCTCGGCTTCAGAACCGAGGCTGAGTCCCAAGGCTGGGCC 71  
Db 312 ATGGCGACTGGCGCGAACCCTCGGCTTCAGAACCGAGGCTGAGTCCCAAGGCTGGGCC 253  
QY 72 AAGGCTGACGAGGAGGAGGATGAGTTAAGGGGCTAGGACCAAGGAGAAAGGTGTATCG 131  
Db 252 AAGGCTGACGAGGAGGAGGATGAGTTAAGGGGCTAGGACCAAGGAGAAAGGTGTATCG 193  
QY 132 CGGGCTGTGGCGCTGCGACATACAAGACCATGGGGCCAGCGTGGGATCAGCAGGAGAA 191  
Db 192 CGGGCTGTGGCGCTGCGACATACAAGACCATGGGGCCAGCGTGGGATCAGCAGGAGAA 133  
QY 192 GCGCTGAGCGAGAGCGATGGGATGAGTACGCCATGGCTTCCTCCGCGAGAGCTCCCC 251  
Db 132 GCGCTGAGCGAGAGCGATGGGATGAGTACGCCATGGCTTCCTCCGCGAGAGCTCCCC 73  
QY 252 GGGGAGTACGAGTGGGAATATGACGAAGAGGAGGAGAGAAACACGCTGGAGATTGAGAGA 311  
Db 72 GGGGAGTACGAGTGGGAATATGACGAAGAGGAGGAGAGAAACACGCTGGAGATTGAGAGA 13  
QY 312 CTGGAGGAGCAG 323  
Db 12 CTGGAGGAGCAG 1

Search completed: June 22, 2003, 07:08:46  
Job time : 535 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2003, 06:58:01 ; Search time 127 seconds  
(without alignments)  
9171.326 Million cell updates/sec

Title: US-09-763-985A-1  
Perfect score: 3798  
Sequence: 1 ccacgcgtccgactggcgact.....tcaaaaaaaaaaaaaaaaaa 3798

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents\_NA:\*

1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq:\*

2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq:\*

3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq:\*

4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq:\*

5: /cgn2\_6/ptodata/1/ina/PCBUS.COMB.seq:\*

6: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
C 1	95.6	2.5	7218	1	US-08-232-463-14
C 2	57.8	1.5	1926	4	US-09-249-385A-4
C 3	57.8	1.5	1931	2	US-09-130-114-2
C 4	51.4	1.4	436	4	US-09-397-787-257
C 5	49.4	1.3	51259	3	US-08-781-891-209
C 6	47.6	1.3	2277	1	US-08-676-967-2
C 7	47.6	1.3	2277	1	US-08-676-974-2
C 8	47.6	1.3	2277	2	US-09-098-487-2
C 9	47.4	1.2	289	4	US-09-007-005-17
C 10	47.4	1.2	289	4	US-09-244-796-17
C 11	47.4	1.2	1926	4	US-09-249-385A-2
C 12	47.4	1.2	2580	3	US-09-050-863-2
C 13	47.4	1.2	2580	4	US-09-359-081-2
C 14	47.4	1.2	5452	2	US-09-130-114-1
C 15	47.4	1.2	9600	4	US-09-910-647-1
C 16	47.4	1.2	9600	4	US-09-620-925-1
C 17	47.4	1.2	10596	1	US-07-884-811-15
C 18	47.4	1.2	10596	1	US-07-885-971-15
C 19	47.4	1.2	10596	1	US-08-087-783A-15
C 20	47.4	1.2	10596	1	US-08-194-088B-15
C 21	47.4	1.2	10596	1	US-08-194-087-15
C 22	47.4	1.2	10596	5	PCT-US93-04648-15
C 23	47.2	1.2	16442	3	US-08-781-891-208
C 24	47	1.2	2338	1	US-08-425-069-1
C 25	47	1.2	2338	2	US-08-317-844B-1
C 26	44.2	1.2	456	2	US-08-557-309B-16
C 27	44.2	1.2	456	3	US-08-834-306-16

28	44.2	1.2	456	4	US-08-993-674A-16
29	44.2	1.2	456	4	US-09-256-976-16
30	44.2	1.2	3489	2	US-08-728-323A-1
31	44.2	1.2	3489	4	US-09-298-568-1
C 32	44.2	1.2	32207	2	US-08-770-379-20
C 33	44.2	1.2	32207	4	US-08-757-669A-20
C 34	44.2	1.2	32207	4	US-09-230-371A-20
35	44	1.2	18443	4	US-09-078-294-6
36	42.4	1.1	282	4	US-09-461-697-205
37	42.4	1.1	306	4	US-09-461-697-203
38	42.4	1.1	696	4	US-09-461-697-193
39	42.4	1.1	699	4	US-09-461-697-191
40	42.4	1.1	717	4	US-09-461-697-189
41	42.4	1.1	774	4	US-09-461-697-187
42	42.4	1.1	819	4	US-09-461-697-185
43	42.4	1.1	1669	4	US-09-461-697-184
44	42.2	1.1	1052	1	US-08-466-603-1
45	42.2	1.1	1052	1	US-08-314-503A-1

ALIGNMENTS

RESULT 1  
US-08-232-463-14/c  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.  
; APPLICANT: FALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-0299  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/232,463  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/935,313  
; FILING DATE:  
; APPLICATION NUMBER: EP 91 114 300.6  
; FILING DATE: 26-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 30472/114 IMM  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)836-9300  
; TELEFAX: (703)683-4109  
; TELEX: 899149  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7218 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; CLONE: pTZgpt-F1s  
US-08-232-463-14

Query Match 2.5%; Score 95.6; DB 1; Length 7218;

[illegible]

```

RESULT 2
US-09-249-585A-4/c
; Sequence 4, Application US/09249585A
; Patent No. 6417002
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert
; TITLE OF INVENTION: METHOD FOR MAINTENANCE AND SELECTION OF EPISODES
; FILE REFERENCE: 0867/00905
; CURRENT APPLICATION NUMBER: US/09/249,585A
; CURRENT FILING DATE: 1999-02-11
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 1926
; TYPE: DNA
; ORGANISM: Epstein Barr Virus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1926)
; OTHER INFORMATION: template strand of EBNA-1 DNA
US-09-249-585A-4

```

Query Match	1.5%	Score	57.8;	DB	4;	Length	1926;
Best Local Similarity	47.0%;	Pred.	No. 7.6e-06;				
Matches	179;	Conservative	0;	Mismatches	202;	Indels	0;
Gaps	0;						
QY	42	GAACCCGAGCGCTGAGTCCAAAGCTGGCGCCCAAGGCTGACGGAGAGGAGGATGAGGTTAAG	101				
Db	913	GAGGACGGGAGGACGGGAGAGACGAGAGCGGGAGGACGGGGAGGACGAGGACGGGGAG	854				
QY	102	GCGGCTAGGACACAGGAGAAGGTGTATCCGGGGCTGTGCGCGCTCGCGACATACAAAGACC	161				
Db	853	GACGGGAGGAGGACGAGGACGGGGAGGACGGGGAGGAGGACGAGGACGGGGAGGACGGG	794				
QY	162	ATGCGGCCACGCTGGGGCTCAGCAGGAGGAAGCGCTGAGCGACGAGCGCATGGGATGAGTAC	221				
Db	793	GAGGACGACGACGGGAGGACCGGGAGGACGAGGACGGGGAGGACCGGGAGGACGAGGAC	734				
QY	222	GCCATGGCTCTTCGCGCGGAGAGCTCCCGCCGGGGAGTACGAGTGGGAATATGACGAAAGAG	281				
Db	733	GGGGAGGACGAGGACGGGAGGACGAGGACGAGGAGGACGGGGAGGACGAGGACGGGGAGGACGGGGAG	674				

RESULT 4  
US-09-397-787-257  
; Sequence 257, Application US/09397787  
; Patent No. 6468758  
; GENERAL INFORMATION:



[illegible]

Qy	133	GGCTGTGGCCGTCGCACATACAAGACCATGGGCCAGCTGGGTATCAGCAGGAGGAAG	192
Db	198	AAGAGGAGGAGGAGGAGAGAAGAAGCAGCAGCAGGAGGAGGAGCAGGAGGAGGAGG	139
Qy	193	GCCTGAGCGAGACGGATGGGGATGAGTACGCCATGCTTCCCGGGAGAGCTCCCCG	252
Db	138	AGAAGGAGGAGNAGGAGGAGGAGAAGGAGGGCAGGAGGAGGAGGAGGAGGAGG	79
Qy	253	GGGAGTCAGTAGTGGGGAATATGACAGAGGAGGAGAAACAGCTGGAGATTGAGAGAC	312
Db	78	AGAGCAGGAGCAGGAACAGCAGCAGGAGGAGGAGGAGGAGCAGGAGGAGCAGCAGGAGG	19
Qy	313	TGGAGGAGCAG	323
Db	18	AGCAGGAGGAG	8

RESULT 6  
US-08-676-967-2  
; Sequence 2, Application US/08676967  
; Patent No. 5747317  
; GENERAL INFORMATION:  
; APPLICANT: COLLINS, KATHLEEN  
; TITLE OF INVENTION: Human Telomerase  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Science & Technology Law Group  
; STREET: 268 Bush Street, Suite 3200  
; CITY: San Francisco  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/676,967  
; FILING DATE:  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Osman Ph.D., Richard A  
; REGISTRATION NUMBER: 36,627  
; REFERENCE/DOCKET NUMBER: UCB96-055  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415)343-4341  
; TELEFAX: (415)343-4342  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2277 base pairs  
; TYPE: nucleic acid

STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cdna  
US-08-676-967-2

Query Match 1.3%; Score 47.6; DB 1; Length 2277;

Best Local Similarity 32.9%; Pred. No. 0.0057;

Matches 74; Conservative 36; Mismatches 115; Indels 0; Gaps 0;

QY 2407 AGAACACAAAGCTGTTTCATCTCAGGCGCTTCCTGCTACTAAAGAGGAACTAGAG 2466

DB 335 ARAARGCNGMNTNATHATHMGNAAAYTNWSNTTYAARTGYWSNGARGAYGAYYTNAARA 394

QY 2467 AAATCTGTAAGGCTCATGCGACCGTGAAGGACCTCAGGCTGCTCACCACCGGCTGGCA 2526

DB 395 CNGTNTTYCCNARTTYGGNGCNGTNYNGARGTNAAYATHCCNMGNARCCNGAYGGNA 454

QY 2527 AACCAAGGCGCTGCTACGTGGAGTATGAAAAATGAATCCAGGCGTGGAGCTGTGA 2586

DB 455 ARATGMGNGNTTYGGNTTYGNCARTTYAARAAYTNTNGARGCNGGNAARGCNYTNA 514

QY 2587 TGAAGATGACGCGCATGACTATCAAGAGAACATCATCAAGTGG 2631

DB 515 ARGGNATGAAYATGAARGARATHAARGNMGNACNGTNGCNGTNG 559

## RESULT 7

US-08-676-974-2

; Sequence 2, Application US/08676974

; Patent No. 5770422

; GENERAL INFORMATION:

; APPLICANT: COLLINS, KATHLEEN

; TITLE OF INVENTION: Human Telomerase

; NUMBER OF SEQUENCES: 10

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Science & Technology Law Group

; STREET: 268 Bush Street, Suite 3200

; CITY: San Francisco

; STATE: CA

; COUNTRY: USA

; ZIP: 94104

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/676,974

; FILING DATE:

; CLASSIFICATION: 530

; ATTORNEY/AGENT INFORMATION:

; NAME: Osman Ph.D., Richard A

; REGISTRATION NUMBER: 36,627

; REFERENCE/DOCKET NUMBER: UCB96-055

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415)343-4341

; TELEFAX: (415)343-4342

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2277 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: cdna

US-08-676-974-2

Query Match

Best Local Similarity 1.3%; Score 47.6; DB 1; Length 2277;

Best Local Similarity 32.9%; Pred. No. 0.0057;

Matches 74; Conservative 36; Mismatches 115; Indels 0; Gaps 0;

QY 2407 AGAACACAAAGCTGTTTCATCTCAGGCGCTTCCTGCTACTAAAGAGGAACTAGAG 2466

DB 335 ARAARGCNGMNTNATHATHMGNAAAYTNWSNTTYAARTGYWSNGARGAYGAYYTNAARA 394

QY 2467 AAATCTGTAAGGCTCATGCGACCGTGAAGGACCTCAGGCTGCTCACCACCGGCTGGCA 2526

DB 395 CNGTNTTYCCNARTTYGGNGCNGTNYNGARGTNAAYATHCCNMGNARCCNGAYGGNA 454

QY 2527 AACCAAGGCGCTGCTACGTGGAGTATGAAAAATGAATCCAGGCGTGGAGCTGTGA 2586

DB 455 ARATGMGNGNTTYGGNTTYGNCARTTYAARAAYTNTNGARGCNGGNAARGCNYTNA 514

QY 2587 TGAAGATGACGCGCATGACTATCAAGAGAACATCATCAAGTGG 2631

DB 515 ARGGNATGAAYATGAARGARATHAARGNMGNACNGTNGCNGTNG 559

## RESULT 8

US-09-098-487-2

; Sequence 2, Application US/09098487

; Patent No. 5917025

; GENERAL INFORMATION:

; APPLICANT: COLLINS, Kathleen

; TITLE OF INVENTION: Human Telomerase

; NUMBER OF SEQUENCES: 11

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Science & Technology Law Group

; STREET: 268 Bush Street, Suite 3200

; CITY: San Francisco

; STATE: CA

; COUNTRY: USA

; ZIP: 94104

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/098,487

; FILING DATE:

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Osman Ph.D., Richard A

; REGISTRATION NUMBER: 36,627

; REFERENCE/DOCKET NUMBER: UCB96-055

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415)343-4341

; TELEFAX: (415)343-4342

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2277 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: cdna

US-09-098-487-2

Query Match

Best Local Similarity 1.3%; Score 47.6; DB 2; Length 2277;

Best Local Similarity 32.9%; Pred. No. 0.0057;

Matches 74; Conservative 36; Mismatches 115; Indels 0; Gaps 0;

QY 2407 AGAACACAAAGCTGTTTCATCTCAGGCGCTTCCTGCTACTAAAGAGGAACTAGAG 2466

DB 335 ARAARGCNGMNTNATHATHMGNAAAYTNWSNTTYAARTGYWSNGARGAYGAYYTNAARA 394

QY 2467 AAATCTGTAAGGCTCATGCGACCGTGAAGGACCTCAGGCTGCTCACCACCGGCTGGCA 2526

DB 395 CNGTNTTYCCNARTTYGGNGCNGTNYNGARGTNAAYATHCCNMGNARCCNGAYGGNA 454

QY 2527 AACCAAGGCGCTGCTACGTGGAGTATGAAAAATGAATCCAGGCGTGGAGCTGTGA 2586

DB 455 ARATGMGNGNTTYGGNTTYGNCARTTYAARAAYTNTNGARGCNGGNAARGCNYTNA 514

QY 2587 TGAAGATGACGCGCATGACTATCAAGAGAACATCATCAAGTGG 2631

DB 515 ARGGNATGAAYATGAARGARATHAARGNMGNACNGTNGCNGTNG 559

```

; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; EARLIER APPLICATION NUMBER: 09/007,005
; EARLIER FILING DATE: 1998-01-14
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 289
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; NAME/KEY: misc_feature
; LOCATION: (1)...(289)
; OTHER INFORMATION: n = A,T,C or G
US-09-244-796-17

Query Match 1.2%; Score 47.4; DB 4; Length 289;
Best Local Similarity 6.9%; Pred. No. 0.0019;
Matches 17; Conservative 11; Mismatches 118; Indels 0; Gaps 0;

QY 1748 GAGAATCAAGGCTGCAGAGAAGGAAGCAGCCCTTGTGCAGCAAGAAGAAAGGCTGA 1807
: : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 25 RARURURARCARARURURARCRAPARURGNRNRNRNRNRNRNRNRNRNRNRN 84

QY 1808 ACAACGAAAGAGCTCGGGCTGAGAGAAGCGTTAAAAAGAAAGATCAGAGG 1867
: : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 85 RSRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRN 144

QY 1868 CCCAGAGAACCGCGGACACATGAGGACGATGAGAAAGACTGGGGCGATGATGAAGA 1927
: : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 145 RSRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRN 204

QY 1928 GCAGCCTTCCAAACGCGAAGGTCGAGAAACAGCATCCCTGCAGCTGGAGAAACACAAA 1987
: : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 205 RSRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRN 264

QY 1988 TGTAGAA 1994
: | |
Db 265 AAAAAAA 271

RESULT 11
US-09-249-585A-2
; Sequence 2, Application US/09249585A
; Patent No. 6417002
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert
; TITLE OF INVENTION: METHOD FOR MAINTENANCE AND SELECTION OF EPISODES
; FILE REFERENCE: 0867/0D905
; CURRENT APPLICATION NUMBER: US/09/249,585A
; CURRENT FILING DATE: 1999-02-11
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 1926
; TYPE: DNA
; ORGANISM: Epstein Barr Virus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1926)
; OTHER INFORMATION: coding strand of EBNA-1 DNA
US-09-249-585A-2

Query Match 1.2%; Score 47.4; DB 4; Length 1926;
Best Local Similarity 43.4%; Pred. No. 0.0059;
Matches 219; Conservative 0; Mismatches 286; Indels 0; Gaps 0;

QY 42 GAACCCGAGGCTGAGTCCAGGCTGGGCCCAAGGCTGACGAGAGGAGGATGAGTTAAG 101
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 527 GAGCAGGAGGAGGGGCGAGGAGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 586

```



```
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-65638/DJB/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 949-8711
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2580 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-359-081-2

Query Match      1.2%; Score 47.4; DB 4; Length 2580;
Best Local Similarity 43.4%; Pred. No. 0.007;
Matches 219; Conservative 0; Mismatches 286; Indels 0; Gaps 0;

QY 42 GAACCCGAGCTGAGTCCAAAGCTGGGCCCAAGCTGACGAGAGGAGGATGAGGTTAAG 101
Db 910 GAGCAGGAGGAGGGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 969
QY 102 GCGGCTAGGACAAAGGAGAAAGTGTATCGCGGGCTGTTCGCGGCTGCGACATACAAAGC 161
Db 970 CAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1029
QY 162 ATGGGGCCAGGCTGGGATCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGTAC 221
Db 1030 GAGGGCCAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1089
QY 222 GCCATGGCTTCTCCCGGAGAGCTCCCGGGGAGTACGAGTGGGATATGACGAAGAG 281
Db 1090 GAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1149
QY 282 GAGGAGAAACACCTGAGATTGAGAGACTGGAGGAGCAGTGTCTATCAACCTCTAT 341
Db 1150 GAGGGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1209
QY 342 GACTACAACTGCCATGTGGATTGATCAGACTGCTCAGGCTGGAAGGGAGCTTACCAAG 401
Db 1210 GAGGGCAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1269
QY 402 GTGAGGATGGCCCGCAGAGATGATGAAATCTTCCCTTCACTGAAGAGCTCTGGCTG 461
Db 1270 GAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1329
QY 462 GAGTGGCTGCATGACGAGATCAGCATGATGCCCGCAGGATGGCTGGACAGAGCAGCTGTAT 521
Db 1330 GGGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1389
QY 522 GACCTCTTTGAGAAAGCCGTGAAGG 546
Db 1390 GCCGGGTCGAGGAGGTAGTGGAGG 1414

RESULT 14
US-09-130-114-1/c
; Sequence 1, Application US/09130114
; Patent No. 5976807
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert A.
; APPLICANT: Damaj, Bassam B.
; APPLICANT: Robbins, Alan K.
; TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
; FILE REFERENCE: 0867/ID903US1
; CURRENT APPLICATION NUMBER: US/09/130,114
; FILING DATE: 1998-08-06
; NUMBER OF SEQ ID NOS: 36
```

```
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 5452
; TYPE: DNA
; ORGANISM: VEBNA
US-09-130-114-1

Query Match      1.2%; Score 47.4; DB 2; Length 5452;
Best Local Similarity 43.4%; Pred. No. 0.011;
Matches 219; Conservative 0; Mismatches 286; Indels 0; Gaps 0;

QY 42 GAACCCGAGGCTGAGTCCAAAGCTGGGCCCAAGCTGACGAGAGGAGGAGGATGAGGTTAAG 101
Db 1895 GAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1836
QY 102 GCGGCTAGGACAAAGGAGAAAGTGTATCGCGGGCTGTTCGCGGCTGCGACATACAAAGC 161
Db 1835 CAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1776
QY 162 ATGGGGCCAGGCTGGGATCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGTAC 221
Db 1775 GAGGGCCAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1716
QY 222 GCCATGGCTTCTCCCGGAGAGCTCCCGGGGAGTACGAGTGGGATATGACGAAGAG 281
Db 1715 GAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1656
QY 282 GAGGAGAAACACCTGAGATTGAGAGACTGGAGGAGCAGTGTCTATCAACCTCTAT 341
Db 1655 GAGGGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1596
QY 342 GACTACAACTGCCATGTGGATTGATCAGACTGCTCAGGCTGGAAGGGAGCTTACCAAG 401
Db 1595 GAGGGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1536
QY 402 GTGAGGATGGCCCGCAGAGATGATGAAATCTTCCCTTCACTGAAGAGCTCTGGCTG 461
Db 1535 GAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1476
QY 462 GAGTGGCTGCATGACGAGATCAGCATGATGCCCGCAGGATGGCTGGACAGAGCAGCTGTAT 521
Db 1475 GGGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1416
QY 522 GACCTCTTTGAGAAAGCCGTGAAGG 546
Db 1415 GCCGGGTCGAGGAGGTAGTGGAGG 1391

RESULT 15
US-08-910-647-1
; Sequence 1, Application US/08910647
; Patent No. 6251433
; GENERAL INFORMATION:
; APPLICANT: Zuckermann et al.
; TITLE OF INVENTION: Compositions and Methods for
; TITLE OF INVENTION: Polynucleotide Delivery
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/910,647
; FILING DATE:
; CLASSIFICATION: 514
```



1	2886	76.0	3537	10	US-09-821-883-30	Sequence 30, Appl
2	2563.2	67.5	3355	10	US-09-925-300-556	Sequence 556, App
3	1056.4	27.8	32145	9	US-09-764-691-7319	Sequence 7319, Ap
4	441.4	11.6	567	9	US-09-918-995-27590	Sequence 27590, A
5	393.4	10.4	432	10	US-09-867-701-1728	Sequence 1728, Ap
6	392.4	10.3	430	10	US-09-867-701-2028	Sequence 2028, Ap
7	310.4	8.2	312	10	US-09-864-761-19458	Sequence 19458, A
8	284	7.5	284	9	US-09-796-692-2624	Sequence 2624, Ap
9	284	7.5	284	9	US-10-040-862-2624	Sequence 2624, Ap
10	244	6.4	244	9	US-09-796-692-6818	Sequence 6818, Ap
11	244	6.4	244	9	US-10-040-862-6818	Sequence 6818, Ap
12	219.8	5.8	475	10	US-09-864-761-2742	Sequence 2742, Ap
13	129.6	3.4	185	10	US-09-864-761-18092	Sequence 18092, A
14	126.4	3.3	467	10	US-09-864-761-1331	Sequence 1331, Ap
15	55.6	1.5	575	10	US-09-864-761-22608	Sequence 22608, A
16	55.4	1.5	572	10	US-09-864-761-20733	Sequence 20733, A
17	55.4	1.5	1282	9	US-10-002-344A-89	Sequence 89, Appl
18	55.4	1.5	1969	10	US-09-864-761-3972	Sequence 3972, Ap
19	54.8	1.4	554	9	US-10-101-487-69	Sequence 69, Appl

181	Db		GGCGTGAGGAGGAGCGATGGGATGAGTACGCCATGGCTTCCTCCGCGGAGAGCTCCGCC	240
252	QY		GGGAGCTACGAGTGGGAATATACGGAAGAGGAGGAGAAAACACCTCGAGATTGAGAGA	311
241	Db		GGGAGTACGAGTGGGAATATACGGAAGAGGAGGAGAAAACACCTCGAGATTGAGAGA	300
312	QY		CTGGAGGACAGTGTCTATCAACGCTCTATGACTACAACTGCCATGTGGACTTGATCAGA	371
301	Db		CTGGAGGACAGTGTCTATCAACGCTCTATGACTACAACTGCCATGTGGACTTGATCAGA	360
372	QY		CTGCTCAGGCTGGAAGGGAGCTTACCAAGGTGAGGATGGCCCGCCAGAAAGATGAGTAA	431
361	Db		CTGCTCAGGCTGGAAGGGAGCTTACCAAGGTGAGGATGGCCCGCCAGAAAGATGAGTAA	420
432	QY		ATCTTTCCCTTGACTGAAGAGCTCTGGCTGGAGTGGCTCATGACGAGATCAGCATGGCC	491
421	Db		ATCTTTCCCTTGACTGAAGAGCTCTGGCTGGAGTGGCTCATGACGAGATCAGCATGGCC	480
492	QY		CAGGATGGCTGGACAGAGACAGCTGTATGACCTCTTTGAGAAAGCCGTGAAGATTAC	551
481	Db		CAGGATGGCTGGACAGAGACAGCTGTATGACCTCTTTGAGAAAGCCGTGAAGATTAC	540
552	QY		ATTTGCTCTAACATTTGGCTAGAGTATGCCAGTACTCATGTTGGTGGGANTGGTCAGAAA	611
541	Db		ATTTGCTCTAACATTTGGCTAGAGTATGCCAGTACTCATGTTGGTGGGANTGGTCAGAAA	600
612	QY		GGTGGCTTTGAGAAAGTTCCGCTCGCTGTTTGAAGAGGCTCTCGTCTGTTGGTTTACAT	671
601	Db		GGTGGCTTTGAGAAAGTTCCGCTCGCTGTTTGAAGAGGCTCTCGTCTGTTGGTTTACAT	660
672	QY		ATGACAAAGGACTCGCCCTCTGGAGGCTTACCAGAGTGTGAAAGTGCATTTGGAA	731
661	Db		ATGACAAAGGACTCGCCCTCTGGAGGCTTACCAGAGTGTGAAAGTGCATTTGGAA	720
732	QY		GCTGCTCGCTTTGAGAAAGTCCACAGTCTTTTCCGGCGACAGTTGGCGATCCACTAT	791
721	Db		GCTGCTCGCTTTGAGAAAGTCCACAGTCTTTTCCGGCGACAGTTGGCGATCCACTAT	780
792	QY		GATATGGAGGCGACATTTGACAGTATGAAGATGGTCAGAGACCCCAATACAGAGTCA	851
781	Db		GATATGGAGGCGACATTTGACAGTATGAAGATGGTCAGAGACCCCAATACAGAGTCA	840
852	QY		GTAATTCAGAACTATAACAAAGCACTACAGCAGCTGGAGAAATATAAACCTTATGAAGA	911
841	Db		GTAATTCAGAACTATAACAAAGCACTACAGCAGCTGGAGAAATATAAACCTTATGAAGA	900
912	QY		GCACCTGCTCAGCAGAGCCACCAAGGCTGGCAGAAATATCAAGCATATATCGATTTGAG	971
901	Db		GCACCTGCTCAGCAGAGCCACCAAGGCTGGCAGAAATATCAAGCATATATCGATTTGAG	960
972	QY		ATGAAAATTTGGCGATCTCTCGCATTCAGTTGATCTTTGAGCGCGCCCTGGTCAGAAC	1031
961	Db		ATGAAAATTTGGCGATCTCTCGCATTCAGTTGATCTTTGAGCGCGCCCTGGTCAGAAC	1020
1032	QY		TGCTTTGCCAGACTATGAGTCCGTTACAGTACAGTACCTAGATGCAAACTGAAAGTA	1091
1021	Db		TGCTTTGCCAGACTATGAGTCCGTTACAGTACAGTACCTAGATGCAAACTGAAAGTA	1080
1092	QY		AAGGATTTGGTTTATCTGATACATACCCGCTATTAGAACTGCCCTGGACATTGCC	1151
1081	Db		AAGGATTTGGTTTATCTGATACATACCCGCTATTAGAACTGCCCTGGACATTGCC	1140
1152	QY		TTATGAGTTCGTTACCTTTGGCCATGGAGAGACATGGAGTTGATCATCAAGTAATTTCT	1211
1141	Db		TTATGAGTTCGTTACCTTTGGCCATGGAGAGACATGGAGTTGATCATCAAGTAATTTCT	1200
1212	QY		GTAACCTTCGAGAAAGCTTTGAATGCCGCTTTCATCCAGGCCACTGATATGTGAGATT	1271
1201	Db		GTAACCTTCGAGAAAGCTTTGAATGCCGCTTTCATCCAGGCCACTGATATGTGAGATT	1260
1272	QY		TGCGAGGATACCTTGATTACCTGAGGAGAGGGTTGATTCAACAAGACTCCAGTAA	1331
1261	Db		TGCGAGGATACCTTGATTACCTGAGGAGAGGGTTGATTCAACAAGACTCCAGTAA	1320
1332	QY		GAGCTGGAGGAGCTTCAGGGCCGCCCTTTACTCTGCTGCTTGGAGTATCTGAAGCAGAGG	1391
1321	Db		GAGCTGGAGGAGCTTCAGGGCCGCCCTTTACTCTGCTGCTTGGAGTATCTGAAGCAGAGG	1380
1392	QY		GAAGAGCGTTTCAATGAGAGTGGTGATCCAAGCTCGCTGATTATGACAGAACTGGCTAGG	1451
1381	Db		GAAGAGCGTTTCAATGAGAGTGGTGATCCAAGCTCGCTGATTATGACAGAACTGGCTAGG	1440
1452	QY		ATTGAGGCTGACTGTGCAATGCAATGCAAGAGCTCGGAACTCTGGGATGAGCATCATG	1511
1441	Db		ATTGAGGCTGACTGTGCAATGCAATGCAAGAGCTCGGAACTCTGGGATGAGCATCATG	1500
1512	QY		ACCAGAGAAATGCGCAAGTACGCCAACATGTGGCTAGAGTATTACAACTGGGAAAGCT	1571
1501	Db		ACCAGAGAAATGCGCAAGTACGCCAACATGTGGCTAGAGTATTACAACTGGGAAAGCT	1560
1572	QY		CATGCTGACACCCAGCACTGCGGAAAGCTCTGCAACCGGCGCTCAGTGCACCACTGAC	1631
1561	Db		CATGCTGACACCCAGCACTGCGGAAAGCTCTGCAACCGGCGCTCAGTGCACCACTGAC	1620
1632	QY		TACCCAGAGCAGCTCTGGAAGTGTACTCACCATTGGAGAGGACAGAGGTTCTTTAGAA	1691
1621	Db		TACCCAGAGCAGCTCTGGAAGTGTACTCACCATTGGAGAGGACAGAGGTTCTTTAGAA	1680
1692	QY		GATTTGGGATATAGCTGTTTCAGAAAACCTGAAACCCGATTAGCTGCTCAATGAGCAGAGA	1751
1681	Db		GATTTGGGATATAGCTGTTTCAGAAAACCTGAAACCCGATTAGCTGCTCAATGAGCAGAGA	1740
1752	QY		ATGAAAGGCTGCAGAGAAAGAGCAGCCCTTTGTCAGCAAGAAAGAAAGGCTCGAACAA	1811
1741	Db		ATGAAAGGCTGCAGAGAAAGAGCAGCCCTTTGTCAGCAAGAAAGAAAGGCTCGAACAA	1800
1812	QY		CGGAAAGAGCTCGGGCTGAGAGAAAGAGGTTTAAAGAAAGAAAGATCAGAGGCCCA	1871
1801	Db		CGGAAAGAGCTCGGGCTGAGAGAAAGAGGTTTAAAGAAAGAAAGATCAGAGGCCCA	1860
1872	QY		GAGAAAGCGGAGCAGATGAGGACCATGAGAAAGAGTGGGCGGATGATGAAGAGAGCAG	1931
1861	Db		GAGAAAGCGGAGCAGATGAGGACCATGAGAAAGAGTGGGCGGATGATGAAGAGAGCAG	1920
1932	QY		CCTTCCAAAGCAGAAAGGCTCGAGAACACATCCCTGCGCTGAGGAAACACAAATGTA	1991
1921	Db		CCTTCCAAAGCAGAAAGGCTCGAGAACACATCCCTGCGCTGAGGAAACACAAATGTA	1980
1992	QY		GAGTACAGCAGGCGCCCTGGGAAATGCTGCTGCTGATAGTGTGAGCCCTTCGAG	2051
1981	Db		GAGTACAGCAGGCGCCCTGGGAAATGCTGCTGCTGATAGTGTGAGCCCTTCGAG	2040
2052	QY		CAGAAAGGAGAGGAGCCCTCCCTGAAAGAGGACATGCCCAAGGCTGCTGCACGACAGCAG	2111
2041	Db		CAGAAAGGAGAGGAGCCCTCCCTGAAAGAGGACATGCCCAAGGCTGCTGCACGACAGCAG	2100
2112	QY		AAGGACAGCATCACCGTCTTTGTCAGAAACCTGCGCTTACAGCATGAGAGCGGACAG	2171
2101	Db		AAGGACAGCATCACCGTCTTTGTCAGAAACCTGCGCTTACAGCATGAGAGCGGACAG	2160
2172	QY		AAGCTCAGGCCACTCTTCGAGCCCTGTCGAGGAGTGGTCCAGATCCGACCATCTTCAGC	2231
2161	Db		AAGCTCAGGCCACTCTTCGAGCCCTGTCGAGGAGTGGTCCAGATCCGACCATCTTCAGC	2220
2232	QY		AACCTGGGGATTTCCGAGGTTTACTGCTACGCTGAGGTTTAAAGAGAGAAATTCAGCCCTT	2291
2221	Db		AACCTGGGGATTTCCGAGGTTTACTGCTACGCTGAGGTTTAAAGAGAGAAATTCAGCCCTT	2280
2292	QY		CAGGACCTGAGAGATGGACCCGGAAGTGTAGAAGGAGGCCCAATGTTTTCGCCCTGT	2351
2281	Db		CAGGACCTGAGAGATGGACCCGGAAGTGTAGAAGGAGGCCCAATGTTTTCGCCCTGT	2340
2352	QY		GTGGATAGAGCAAAACCCCGATTTTAAAGTGTTCAGTACAGCATTCCCTAGAGAAA	2411
2341	Db		GTGGATAGAGCAAAACCCCGATTTTAAAGTGTTCAGTACAGCATTCCCTAGAGAAA	2400









Db 121 AGCGAAACGAGAACACACCTGTTTGTGAGACACAGTTGTCCAATGTTTCTGCCAGCTC 180  
QY 3242 CGGCCCTTTTCTATGACACTTCTCTCCACCCTGCACAGACATGTCGCCGTCAATCT 3301  
Db 181 CGGCCCTTTTCTATGACACTTCTCTCCACCCTGCACAGACATGTCGCCGTCAATCT 240  
QY 3302 TTTAATTTAAAGATGAATGGCAGATGCTAGTAATTCACAGATGGGCTCTGTGGGG 3361  
Db 241 TTTAATTTAAAGATGAATGGCAGATGCTAGTAATTCACAGATGGGCTCTGTGGGG 300  
QY 3362 GTGGGTCTGAGGAATGACCTATATAAAACATTTGCTGGAGTTTGTTCATGGGGCTGTG 3421  
Db 301 GTGGGTCTGAGGAATGACCTATATAAAACATTTGCTGGAGTTTGTTCATGGGGCTGTG 360  
QY 3422 CATTATTT-ATATTATGTTGTTGTAATGACATGCTGACCCCTGTTTCATGTTTCCATAAA 3480  
Db 361 CATTATTTAATATTATGTTGTTGTAATGACATGCTGACCCCTGTTTCATGTTTCCATAAA 420  
QY 3481 GCAGATATTTG 3492  
Db 421 GCAGATATTTG 432

RESULT 6

US-09-867-701-2028  
; Sequence 2028, Application US/09867701  
; Patent No. US20020132237A1  
; GENERAL INFORMATION:  
; APPLICANT: Aglate, Paul A.  
; APPLICANT: Jones, Robert  
; APPLICANT: Harlocker, Susan L.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; FILE REFERENCE: 210121.497  
; CURRENT APPLICATION NUMBER: US/09/867,701  
; NUMBER OF SEQ ID NOS: 10912  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2028  
; LENGTH: 430.  
; TYPE: DNA  
; ORGANISM: Homo sapien  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(430)  
; OTHER INFORMATION: n = A,T,C or G

Query Match 10.3%; Score 392.4; DB 10; Length 430;  
Best Local Similarity 98.8%; Pred. No. 3.5e-110;  
Matches 426; Conservative 0; Mismatches 2; Indels 3; Gaps 3;

QY 3064 AAGAGCCTTTCAGTCTCCTCATATTTAGGCGAGTGGCAGAAAAGTGACCACTCTGCAG 3123  
Db 1 AAGAGCCTTTCAGTCTCCTCATATTTAGGCGAGTGGCAGAAAAGTGACCACTCTGCAG 60  
QY 3124 GCTGGGCCAGGATGTGGTCTCCTGAGATAGTTTGTATCTTAAAGACTGAGGCACAGAA 3183  
Db 61 GCTGGG-CCAGATGTGGTCTCCTGAGATAGTTTGTATCTTAAAGACTGAGGCACAGAA 119  
QY 3184 GCGAAACGAGAACACA-CGTGTTTTTGTAGACAGTTGTCCAAATGTTTCTGCCAGCTCC 3242  
Db 120 GCGAAACGAGAACACACCTGTTTTGTAGACACAGTTGTCCAAATGTTTCTGCCAGCTCC 179  
QY 3243 GCGCCCTTTTCTATGACACTTCTCTCCACCCTGCACAGACATGTCGCCCTCATCTT 3302  
Db 180 GCGCCCTTTTGTATGACACTTCTCTCTCCACCCTGCACAGACATGTCGCCCTCATCTT 239  
QY 3303 TTAATTTTAAAGATGAATGGCAGATGCTAGTAATTCACAGATGGGCTCTGTGGGG 3362  
Db 240 TTAATTTTAAAGATGAATGGCAGATGCTAGTAATTCACAGATGGGCTCTGTGGGG 299

QY 3363 TGGGTCTGAGGAAGTCAGCTATATAAACATTTTCTGGAGTTTGTTCATGGGGCTGTGC 3422  
Db 300 TGGGTCTGAGGAAGTCAGCTATATAAACATTTTCTGGAGTTTGTTCATGGGGCTGTGC 359  
QY 3423 ATTTT-ATATTATGTTTGTAAATGACATGTCAGCCCTGTTTTCATGTTTCCATAAAG 3481  
Db 360 ATTTTAAATATTATGTTTGTAAATGACATGTCAGCCCTGTTTTCATGTTTCCATAAAG 419  
QY 3482 CAGATATTTG 3492  
Db 420 CAGATATTTG 430

RESULT 7

US-09-864-761-19458/C  
; Sequence 19458, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO  
; FILE REFERENCE: Aecomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; PRIOR FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 19458  
; LENGTH: 312  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC008119.6  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.1  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.9

OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.8  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.8  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.2  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.3  
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.6  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.5  
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.5  
OTHER INFORMATION: SWISSPROT HIT: P27476, EVALUE 1.10e+00  
OTHER INFORMATION: NT HIT: D63879.1, EVALUE 0.00e+00  
OTHER INFORMATION: EST\_HUMAN HIT: BF514110.1, EVALUE 0.00e+00  
US-09-864-761-19458

Query Match 8.2%; Score 310.4; DB 10; Length 312;  
Best Local Similarity 99.7%; Pred. No. 5.8e-85;  
Matches 311; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 12 ATGGCGACTGCGCGCGAAACCTCGCTTCAGAACCCGAGGTGAGTCAAGGCTGGGCC 71  
Db 312 ATGGCGACTGCGCGCGAAACCTCGCTTCAGAACCCGAGGTGAGTCAAGGCTGGGCC 253  
QY 72 AAGGCTGACGAGAGGAGGATGAGTTAAGCGGCTAGCAAGGAGAAAGGTGTATCG 131  
Db 252 AAGGCTGACGAGAGGAGGATGAGTTAAGCGGCTAGCAAGGAGAAAGGTGTATCG 193  
QY 132 CGGCTGTGGCGCTGCGACATACAAAGACCATGGGCGGCGGATGAGGATGAGGAA 191  
Db 192 CGGCTGTGGCGCTGCGACATACAAAGACCATGGGCGGCGGATGAGGATGAGGAA 133  
QY 192 GCGCTGACGAGAGGCGATGGGATGAGTACGCCATGGCTTCCTCCGCGGAGAGCTCCGCC 251  
Db 132 GCGCTGACGAGAGGCGATGGGATGAGTACGCCATGGCTTCCTCCGCGGAGAGCTCCGCC 73  
QY 252 GGGGAGTACGAGTGGGAATATGACCAAGAGGAGGAGAAACCCAGCTGAGATTGAGAGA 311  
Db 72 GGGGAGTACGAGTGGGAATATGACCAAGAGGAGGAGGAGAAACCCAGCTGAGATTGAGAGA 13  
QY 312 CTGGAGGAGCAG 323  
Db 12 CTGGAGGAGCAG 1

RESULT 8  
US-09-796-692-2624/c  
Sequence 2624, Application US/09796692  
Publication No. US20020198362A1  
GENERAL INFORMATION:  
APPLICANT: Gaiger, Alexander  
APPLICANT: Mannion, Jane  
APPLICANT: Algate, Paul A.  
APPLICANT: Mannion, Marc  
APPLICANT: Corixa Corporation  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY  
FILE REFERENCE: 2077.001200  
CURRENT APPLICATION NUMBER: US/09/796,692  
CURRENT FILING DATE: 2001-03-01  
PRIOR APPLICATION NUMBER: 60/186,126  
PRIOR FILING DATE: 2000-03-01  
PRIOR APPLICATION NUMBER: 60/190,479  
PRIOR FILING DATE: 2000-03-17  
PRIOR APPLICATION NUMBER: 60/200,545  
PRIOR FILING DATE: 2000-04-27  
PRIOR APPLICATION NUMBER: 60/200,303  
PRIOR FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: 60/200,779  
PRIOR FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: 60/200,999  
PRIOR FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 60/202,084  
PRIOR FILING DATE: 2000-05-04  
PRIOR APPLICATION NUMBER: 60/206,201  
PRIOR FILING DATE: 2000-05-22  
PRIOR APPLICATION NUMBER: 60/218,950  
PRIOR FILING DATE: 2000-07-14

PRIOR APPLICATION NUMBER: 60/222,903  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: 60/223,416  
PRIOR FILING DATE: 2000-08-04  
PRIOR APPLICATION NUMBER: 60/223,378  
PRIOR FILING DATE: 2000-08-07  
NUMBER OF SEQ ID NOS: 9597  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 2624  
LENGTH: 284  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-796-692-2624  
Query Match 7.5%; Score 284; DB 9; Length 284;  
Best Local Similarity 100.0%; Pred. No. 7.7e-77;  
Matches 284; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 297 CTGGAGATTGAGAGACTGGAGGAGCAGTTGTCTATCAAGCTCTATGACTACAACTGCCAT 356  
Db 284 CTGGAGATTGAGAGACTGGAGGAGCAGTTGTCTATCAAGCTCTATGACTACAACTGCCAT 225  
QY 357 GTGGACTTGATCAGACTGCTCAGGCTGGAAGGGGAGCTTACCAAGGTGAGGATGGCCGC 416  
Db 224 GTGGACTTGATCAGACTGCTCAGGCTGGAAGGGGAGCTTACCAAGGTGAGGATGGCCGC 165  
QY 417 CAGAAGATGAGTGAATCTTTCCCTTGACTGAAGAGCTCTGGCTGGAGTGGCTGCATGAC 476  
Db 164 CAGAAGATGAGTGAATCTTTCCCTTGACTGAAGAGCTCTGGCTGGAGTGGCTGCATGAC 105  
QY 477 GAGATCAGCATGCGCCAGGATGGCTGGACAGAGCAGCTGTATGACCTCTTTTGAGAAA 536  
Db 104 GAGATCAGCATGCGCCAGGATGGCTGGACAGAGCAGCTGTATGACCTCTTTTGAGAAA 45  
QY 537 GCCGTGAGGATTACATTTGCTTACATTTGGCTAGAGTATGG 580  
Db 44 GCCGTGAGGATTACATTTGCTTACATTTGGCTAGAGTATGG 1

RESULT 9  
US-10-040-862-2624/c  
Sequence 2624, Application US/10040862  
Publication No. US20030078396A1  
GENERAL INFORMATION:  
APPLICANT: Gaiger, Alexander  
APPLICANT: Mannion, Jane  
APPLICANT: Algate, Paul A.  
APPLICANT: Mannion, Marc  
APPLICANT: Corixa Corporation  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY  
FILE REFERENCE: 014058-013520US  
CURRENT APPLICATION NUMBER: US/10/040,862  
CURRENT FILING DATE: 2001-11-06  
PRIOR APPLICATION NUMBER: US 60/186,126  
PRIOR FILING DATE: 2000-03-01  
PRIOR APPLICATION NUMBER: US 60/190,479  
PRIOR FILING DATE: 2000-03-17  
PRIOR APPLICATION NUMBER: US 60/200,545  
PRIOR FILING DATE: 2000-04-27  
PRIOR APPLICATION NUMBER: US 60/200,303  
PRIOR FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: US 60/200,779  
PRIOR FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: US 60/200,999  
PRIOR FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: US 60/202,084  
PRIOR FILING DATE: 2000-05-04  
PRIOR APPLICATION NUMBER: US 60/206,201  
PRIOR FILING DATE: 2000-05-22  
PRIOR APPLICATION NUMBER: US 60/218,950  
PRIOR FILING DATE: 2000-07-14  
PRIOR APPLICATION NUMBER: US 60/222,903

;  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: US 60/223,416  
; PRIOR FILING DATE: 2000-08-04  
; PRIOR APPLICATION NUMBER: US 60/223,378  
; PRIOR FILING DATE: 2000-08-07  
; PRIOR APPLICATION NUMBER: US 09/796,692  
; PRIOR FILING DATE: 2001-03-01  
; NUMBER OF SEQ ID NOS: 10467  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2624  
; LENGTH: 284  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; US-10-040-862-6264

Query Match 7.5%; Score 284; DB 9; Length 284;  
Best Local Similarity 100.0%; Pred. No. 7.7e-77;  
Matches 284; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 297 CTGGAGATTGAGAGACTGGAGGAGCAGTTGCTCTATCAAGCTCTATGACTACAACTGCCAT 356  
Db 284 CTGGAGATTGAGAGACTGGAGGAGCAGTTGCTCTATCAAGCTCTATGACTACAACTGCCAT 225  
QY 357 GTGGACTTGTACAGACTGCTCAGGCTGGAGGGAGCTTACCAAGTGGAGGATGGCCCGC 416  
Db 224 GTGGACTTGTACAGACTGCTCAGGCTGGAGGGAGCTTACCAAGTGGAGGATGGCCCGC 165  
QY 417 CAGAAGTGAAGTGAAGTCTTCCCTTGACTGAAGAGCTCTGGCTGGAGTGGCTGCATGAC 476  
Db 164 CAGAAGTGAAGTGAAGTCTTCCCTTGACTGAAGAGCTCTGGCTGGAGTGGCTGCATGAC 105  
QY 477 GAGATCAGCATGCCCGCAGGATGCCCTGGACAGAGACAGCTGTATGACCTCTTTTGAGAA 536  
Db 104 GAGATCAGCATGCCCGCAGGATGCCCTGGACAGAGACAGCTGTATGACCTCTTTTGAGAA 45  
QY 537 GCCGTGAAGATTACATTTTCTCTACATTTGGCTAGATGAG 580  
Db 44 GCCGTGAAGATTACATTTTCTCTACATTTGGCTAGATGAG 1

RESULT 10  
US-09-796-692-6818/c  
; Sequence 6818, Application US/09796692  
; Publication No. US20020198362A1  
; GENERAL INFORMATION:  
; APPLICANT: Gaiger, Alexander  
; APPLICANT: Mannion, Jane  
; APPLICANT: Retter, Marc  
; APPLICANT: Corixa Corporation  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY  
; FILE REFERENCE: 2077.001200  
; CURRENT FILING DATE: 2001-03-01  
; PRIOR APPLICATION NUMBER: US/09/796,692  
; PRIOR FILING DATE: 2000-03-01  
; PRIOR APPLICATION NUMBER: 60/186,126  
; PRIOR FILING DATE: 2000-03-01  
; PRIOR APPLICATION NUMBER: 60/190,479  
; PRIOR FILING DATE: 2000-03-17  
; PRIOR APPLICATION NUMBER: 60/200,545  
; PRIOR FILING DATE: 2000-04-27  
; PRIOR APPLICATION NUMBER: 60/200,303  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: 60/200,779  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: 60/200,999  
; PRIOR FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 60/202,084  
; PRIOR FILING DATE: 2000-05-04  
; PRIOR APPLICATION NUMBER: 60/206,201  
; PRIOR FILING DATE: 2000-05-22  
; PRIOR APPLICATION NUMBER: 60/218,950  
; PRIOR FILING DATE: 2000-07-14  
; PRIOR APPLICATION NUMBER: 60/222,903  
; PRIOR FILING DATE: 2000-08-03

;  
; PRIOR APPLICATION NUMBER: 60/223,416  
; PRIOR FILING DATE: 2000-08-04  
; PRIOR APPLICATION NUMBER: 60/223,378  
; PRIOR FILING DATE: 2000-08-07  
; NUMBER OF SEQ ID NOS: 9597  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 6818  
; LENGTH: 244  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; US-09-796-692-6818

Query Match 6.4%; Score 244; DB 9; Length 244;  
Best Local Similarity 100.0%; Pred. No. 1.5e-64;  
Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2991 CTTGGTGTCTGCTGCTCGCAACACAGATGGCTCCTCGGCTTTAGACAGAAAGGGAA 3050  
Db 244 CTTGGTGTCTGCTGCTCGCAACACAGATGGCTCCTCGGCTTTAGACAGAAAGGGAA 185  
QY 3051 GGGTTTCTAAGTCAAGAGCTTTTCAGTGTCTCCCTCATATTGAGGCGAGTGGCAGAAAAGT 3110  
Db 184 GGGTTTCTAAGTCAAGAGCTTTTCAGTGTCTCCCTCATATTGAGGCGAGTGGCAGAAAAGT 125  
QY 3111 GACCACTCTCAGGCTGGGCCAGGATGTGGTGTCTCTGAGATAGTTTTGTATCTTAAAGA 3170  
Db 124 GACCACTCTCAGGCTGGGCCAGGATGTGGTGTCTCTGAGATAGTTTTGTATCTTAAAGA 65  
QY 3171 CTGAGGCACAGAGCGAAACGAGACACACTGTTTTTGGACACACACTTGTCCAAATGTTT 3230  
Db 64 CTGAGGCACAGAGCGAAACGAGACACACTGTTTTTGGACACACACTTGTCCAAATGTTT 5  
QY 3231 CTGG 3234  
Db 4 CTGG 1

RESULT 11  
US-10-040-862-6818/c  
; Sequence 6818, Application US/10040862  
; Publication No. US20030078396A1  
; GENERAL INFORMATION:  
; APPLICANT: Gaiger, Alexander  
; APPLICANT: Algate, Paul A.  
; APPLICANT: Mannion, Jane  
; APPLICANT: Retter, Marc  
; APPLICANT: Corixa Corporation  
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Ther  
; FILE REFERENCE: 014058-013520US  
; CURRENT FILING DATE: 2001-11-06  
; PRIOR APPLICATION NUMBER: US/10/040,862  
; PRIOR FILING DATE: 2000-03-01  
; PRIOR APPLICATION NUMBER: US 60/186,126  
; PRIOR FILING DATE: 2000-03-17  
; PRIOR APPLICATION NUMBER: US 60/190,479  
; PRIOR FILING DATE: 2000-03-17  
; PRIOR APPLICATION NUMBER: US 60/200,545  
; PRIOR FILING DATE: 2000-04-27  
; PRIOR APPLICATION NUMBER: US 60/200,303  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: US 60/200,779  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: US 60/200,999  
; PRIOR FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: US 60/202,084  
; PRIOR FILING DATE: 2000-05-04  
; PRIOR APPLICATION NUMBER: US 60/206,201  
; PRIOR FILING DATE: 2000-05-22  
; PRIOR APPLICATION NUMBER: US 60/218,950  
; PRIOR FILING DATE: 2000-07-14  
; PRIOR APPLICATION NUMBER: US 60/222,903  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: US 60/223,416



```

: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: US 09/632,366
: PRIOR FILING DATE: 2000-08-03
: PRIOR APPLICATION NUMBER: GB 24263.6
: PRIOR FILING DATE: 2000-10-04
: PRIOR APPLICATION NUMBER: US 60/236,359
: PRIOR FILING DATE: 2000-09-27
: PRIOR APPLICATION NUMBER: PCT/US01/006656
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00667
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00664
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00669
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00665
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00668
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00663
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00662
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00661
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00670
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: US 60/234,687
: PRIOR FILING DATE: 2000-09-21
: PRIOR APPLICATION NUMBER: US 09/608,408
: PRIOR FILING DATE: 2000-06-30
: PRIOR APPLICATION NUMBER: US 09/774,203
: PRIOR FILING DATE: 2001-01-29
: NUMBER OF SEQ ID NOS: 49117
: SOFTWARE: Annonax Sequence Listing Engine
: SEQ ID NO 18092
: LENGTH: 185
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: OTHER INFORMATION: MAP TO AC008119.6
: OTHER INFORMATION: EXPRESSED IN ADULT L
: OTHER INFORMATION: EXPRESSED IN BRAIN,
: OTHER INFORMATION: EXPRESSED IN PLACENT
: OTHER INFORMATION: EXPRESSED IN LUNG, S
: OTHER INFORMATION: EXPRESSED IN BONE MAR
: OTHER INFORMATION: EXPRESSED IN FETAL L
: OTHER INFORMATION: EXPRESSED IN BT474,
: OTHER INFORMATION: EXPRESSED IN HEART,
: OTHER INFORMATION: EXPRESSED IN HELA, S
: OTHER INFORMATION: EXPRESSED IN HBL100,
: OTHER INFORMATION: EST_HUMAN HIT: BE739
: OTHER INFORMATION: SWISSPROT HIT: P0320
: OTHER INFORMATION: NT HIT: g17661951, E
: US-09-864-761-18092

```

[illegible]

```

Db          4 ACTG 1

RESULT 14
US-09-864-761-1331/C
; Sequence 1331, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLE
; FILE REFERENCE: Aemica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/006666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006657
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006659
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
; SEQ ID NO 1331
; LENGTH: 467
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC008119.6
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN HEPA, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.9
; US-09-864-761-1331

```



Query Match	3.3%;	Score 126.4;	DB 10;	Length 467;
Best local Similarity	99.2%;	Pred. No. 5.3e-28;		
Matches 127;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;
Qy	792	GATATGAGGCCACATTTCGAGAGTATGAAGAATGGTCAGAGACCCCAATACCAAGTCA	851	
Db	444	GATATGAGGCCACATTTCGAGAGTATGAAGAATGGTCAGAGACCCCAATACCAAGTCA	385	
Qy	852	GTAATTCAGAACTATAACAAAGCACTACAGCAGCTGGAGAAATATAAACCCCTATGAAGAA	911	
Db	384	GTAATTCAGAACTATAACAAAGCACTACAGCAGCTGGAGAAATATAAACCCCTATGAAGAA	325	
Qy	912	GCACCTGTT	919	
Db	324	GCACCTGTT	317	

```

RESULT 15
US-09-864-761-22608/c
; Sequence 22608, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aemica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 22608
; LENGTH: 272
; TYPE: DNA

```

**THIS PAGE BLANK (USPTO)**

Result No.	Score	Query Match	Length	ID	Description
1	1844.4	48.6	1865	11 BC032601	BC032601 Homo sapi
2	1752	46.1	1947	11 BC004547	BC004547 Homo sapi
3	1750.2	46.1	1938	11 BC024279	BC024279 Homo sapi
4	867.2	22.8	872	14 BQ231651	BQ231651 AGENCOURT
5	839.6	22.1	920	14 BQ430825	BQ430825 AGENCOURT
6	831.8	21.9	844	14 BQ223392	BQ223392 AGENCOURT

Contact: nisc.mgc@nih.gov

Akhter, N., Ayelle, K., Beckstrom-Sternberg, S.M., Benjamin, B.,  
Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,  
Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,  
Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R.,  
Maduro, Q.L., Masiello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,  
McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W.,  
Tsurgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,  
Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAC Plate: 69 Row: k Column: 6  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 7661951  
This clone has the following problem: frame shifted.

## FEATURES

## source

1..1865  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5534381"  
/tissue="uterus, leiomyosarcoma"  
/clone\_lib="NIH\_MGC\_71"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6"  
523 a 392 c 557 g 393 t

Query Match 48.6%; Score 1844.4; DB 11; Length 1865;

Best Local Similarity 99.7%; Pred. No. 0; Mismatches 6; Indels 0; Gaps 0;  
Matches 1848; Conservative 0;

QY	5	GGGTCCGATGGGACTCGCGCGAAACCTTCGGCTTCAGAACCCGAGGCTGAGTCCAAAGGC	64
DB	12	GCACAAGATGGGACTCGCGCGAAACCTTCGGCTTCAGAACCCGAGGCTGAGTCCAAAGGC	71
QY	65	TGGGCCAAGGCTGACGAGAGAGAGATGAGTTAAGCGCGCTAGACAAAGAGAAAGT	124
DB	72	TGGGCCAAGGCTGACGAGAGAGAGATGAGTTAAGCGCGCTAGACAAAGAGAAAGT	131
QY	125	GTATCCGCGGCTGTGGCGGCTGCGACATACAAAGACCATGGCGCCAGCTGGGATCAGCA	184
DB	132	GTATCCGCGGCTGTGGCGGCTGCGACATACAAAGACCATGGCGCCAGCTGGGATCAGCA	191
QY	185	GGAGGAAGCGGTGAGCGAGAGCGATGGGATGAGTACGCCATGGCTTCCTCCGCGGAGAG	244
DB	192	GGAGGAAGCGGTGAGCGAGAGCGATGGGATGAGTACGCCATGGCTTCCTCCGCGGAGAG	251
QY	245	CTCCCCGGGAGTACGAGTGGGATATGACGAAGAGGAGGAGAAACCCAGCTGGAGAT	304
DB	252	CTCCCCGGGAGTACGAGTGGGATATGACGAAGAGGAGGAGAAACCCAGCTGGAGAT	311
QY	305	TGAGAGACTGGAGGAGCAGTTCTATCAACGCTATGACTACAACTGCGCATGTGGACTT	364
DB	312	TGAGAGACTGGAGGAGCAGTTCTATCAACGCTATGACTACAACTGCGCATGTGGACTT	371
QY	365	GATCAGACTGCTCAGGCTGGAAGGGAGCTTACCAAGGTGAGGATGGCCCGCCAGAAAGT	424
DB	372	GATCAGACTGCTCAGGCTGGAAGGGAGCTTACCAAGGTGAGGATGGCCCGCCAGAAAGT	431
QY	425	GAGTGAATCTTCCCTTGACTGAAGAGCTCTGGCTGGAGTGGCTGCATGACGAGATCAG	484
DB	432	GAGTGAATCTTCCCTTGACTGAAGAGCTCTGGCTGGAGTGGCTGCATGACGAGATCAG	491
QY	485	CATGGCCCAAGGATGGCTGGACAGAGACACCTGTATGACCTCTTTGAGAAAGCGGTGAA	544
DB	492	CATGGCCCAAGGATGGCTGGACAGAGACACCTGTATGACCTCTTTGAGAAAGCGGTGAA	551
QY	545	GGATTACATTTCTCCTAACATTTGGCTAGATGATGGCCAGTACTCAGTTGGTGGGATGG	604
DB	552	GGATTACATTTCTCCTAACATTTGGCTAGATGATGGCCAGTACTCAGTTGGTGGGATGG	611
QY	605	TCAGAAAGGTGGCCCTTGAGAAAGTTCGCTCCGCTGTTTGAAGGGCTCTCTGCTCTTGG	664

DB	612	TCAGAAAGGTGGCCCTTGAGAAAGTTCGCTCCGCTGTTTGAAGGGCTCTCTCGTCTGTGG	671
QY	665	TTTACATATGACCAAGGACTCGCCCTCTGGAGGCTTACCAGAGAGTTTGAAGAGTCGAT	724
DB	672	TTTACATATGACCAAGGACTCGCCCTCTGGAGGCTTACCAGAGAGTTTGAAGAGTCGAT	731
QY	725	TGTGAAGTGTCTGGCTTGAGAAAGTCCACAGTCTTTTCCGCGCAGACAGTTGGCGATCC	784
DB	732	TGTGAAGTGTCTGGCTTGAGAAAGTCCACAGTCTTTTCCGCGCAGACAGTTGGCGATCC	791
QY	785	ACTCTATGATGAGGAGCCACATTTGCGAGATGATGAAGATGGTCAGAGACCCCAATACC	844
DB	792	ACTCTATGATGAGGAGCCACATTTGCGAGATGATGAAGATGGTCAGAGACCCCAATACC	851
QY	845	ACAGTCAGTAATTCAGAACTATAACAAAGCACTACAGCAGCTGGAGAAATATAACCCCTA	904
DB	852	AGAGTCAGTAATTCAGAACTATAACAAAGCACTACAGCAGCTGGAGAAATATAACCCCTA	911
QY	905	TGAAGAAGCACTGTTTCAGGAGAGAGCAACCAAGGCTGGCAGAAATATCAAGCATATATCA	964
DB	912	TGAAGAAGCACTGTTTCAGGAGAGAGCAACCAAGGCTGGCAGAAATATCAAGCATATATCA	971
QY	965	TTTTGAGATGAAATTTGGCGATCTCTGCGCATTCAGTTGATCTTTGAGCGCGCCCTGGT	1024
DB	972	TTTTGAGATGAAATTTGGCGATCTCTGCGCATTCAGTTGATCTTTGAGCGCGCCCTGGT	1031
QY	1025	CCAGAACTGCCCTTGTCCAGACTTATGGATCCGTTACAGTCACTACCTAGATCGACAACCT	1084
DB	1032	CCAGAACTGCCCTTGTCCAGACTTATGGATCCGTTACAGTCACTACCTAGATCGACAACCT	1091
QY	1085	GAAAGTAAAGGATTTGGTTTATCTGTACATAACCGCGCTATTAGAAAGTCCGCTGGAC	1144
DB	1092	GAAAGTAAAGGATTTGGTTTATCTGTACATAACCGCGCTATTAGAAAGTCCGCTGGAC	1151
QY	1145	AGTTGCCCTTATGGAGTCGGTACCTCTTGGCCATGAGAGACATGAGATGATCATCAAGT	1204
DB	1152	AGTTGCCCTTATGGAGTCGGTACCTCTTGGCCATGAGAGACATGAGATGATCATCAAGT	1211
QY	1205	AATTTCTGTAACTTCGAGAAAGCTTTGAATCCCGGCTTCATCCAGGCCACCTGATATGT	1264
DB	1212	AATTTCTGTAACTTCGAGAAAGCTTTGAATCCCGGCTTCATCCAGGCCACCTGATATGT	1271
QY	1265	GGAGATTTGGCAGGACATACCTTTGATTAACCTGAGGAGAGGGTTGATTTCAAAACAAGCTC	1324
DB	1272	GGAGATTTGGCAGGACATACCTTTGATTAACCTGAGGAGAGGGTTGATTTCAAAACAAGCTC	1331
QY	1325	CAGTAAGAGCTGGAGGAGTTGAGGCGCCCTTTACTCGTCTGGAGTATCTGAAGCA	1384
DB	1332	CAGTAAGAGCTGGAGGAGTTGAGGCGCCCTTTACTCGTCTGGAGTATCTGAAGCA	1391
QY	1385	GGAGTGAAGAGCTGTTCAATGAGAGTGGTGCATCCAAAGCTCGCTGATTCAGAACTG	1444
DB	1392	GGAGTGAAGAGCTGTTCAATGAGAGTGGTGCATCCAAAGCTCGCTGATTCAGAACTG	1451
QY	1445	GGCTAGGATTTAGGCTGACCTGCAATAACATGAGAAAGCTCGGGAAGTCTGGGATAG	1504
DB	1452	GGCTAGGATTTAGGCTGACCTGCAATAACATGAGAAAGCTCGGGAAGTCTGGGATAG	1511
QY	1505	CATCATGACGAGAGAAATGCCAAGTACCCAAAGCTGCGCTAGAGTATTAACAACCTGA	1564
DB	1512	CATCATGACGAGAGAAATGCCAAGTACCCAAAGCTGCGCTAGAGTATTAACAACCTGA	1571
QY	1565	AAGAGCTCATGGTACACCCAGCACTGCGGGAAGCTCTGACCGGGCGCTCCAGTGGAC	1624
DB	1572	AAGAGCTCATGGTACACCCAGCACTGCGGGAAGCTCTGACCGGGCGCTCCAGTGGAC	1631
QY	1625	CAGTCACTACCCAGAGCAGCTGCGAAGTGTACTCACCATGGAGAGACAGAAGTTTC	1684
DB	1632	CAGTCACTACCCAGAGCAGCTGCGAAGTGTACTCACCATGGAGAGACAGAAGTTTC	1691
QY	1685	TTTACAAGATTTGGGATATAGCTGTTCAAGAACTCAACCCGATAGCTCGTGTCAATGA	1744

Db	1692	TTTAGAGAGATGGGATATAGCTGTTTCAGAAACATGAAACCCGATATAGCTCGTGTCAATGA	1751
Qy	1745	GCAGAGAATGAAGGCTGCAGAGAAAGAGCAGCCCTTGTGCAGCAGAAAGAAAAGGC	1804
Db	1752	GCAGAGAATGAAGGCTGCAGAGAAAGAGCAGCCCTTGTGCAGCAGAAAGAAAAGGC	1811
Qy	1805	TGAACACCGAAAGAGCTCGGGCTGAGAGAAAGAGCGTTAAAAAAGAAAGAAAA	1858
Db	1812	TGAACACCGAAAGAGCTCGGGCTGAGAGAAAGAGCGTTAAAAAAGAAAGAAAA	1865
LOCUS	BC004547		
DEFINITION	Homo sapiens, Similar to KIAA0156 gene product, clone IMAGE:3958574, mRNA.	1947 bp	mRNA linear HTC 12-JUL-2001
ACCESSION	BC004547		
VERSION	BC004547.1	GI:14709196	
KEYWORDS	HTC.		
SOURCE	Homo sapiens.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 1947)		
TITLE	Strausberg,R.		
JOURNAL	Direct Submission		
REMARK	Submitted (14-MAR-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA		
COMMENT	NIH-MGC Project URL: <a href="http://mgc.nci.nih.gov">http://mgc.nci.nih.gov</a> Contact: MGC help desk Email: <a href="mailto:cgabps-remail.nih.gov">cgabps-remail.nih.gov</a> Tissue procurement: ATCC CDNA Library Preparation: Rubin.Laboratory CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland; Web site: <a href="http://www.nisc.nih.gov/">http://www.nisc.nih.gov/</a> Contact: <a href="mailto:nisc.mgc@nih.gov">nisc.mgc@nih.gov</a> Shevchenko,Y., Wetherby,K.D., Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S., Dietrich,N.L., Guan,X., Gupta,J., Ho,S.-L., Karlins,E., Legaspi,R., Lim,M., Maduro,Q.L., Masllo,C., Mastrian,S.D., Mcclloskey,J.C., McDowell,J., Pearson,R., Snyder,B., Stantripop,S., Thomas,P.J., Tongson,E.E., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A., Zhang,L.-H. and Green,E.D.		
FEATURES	source		
BASE COUNT	546 a	414 c	572 g 415 t
ORIGIN			
Query Match	46.1%	Score 1752;	DB 11; Length 1947;
Best Local Similarity	95.3%	Pred. No. 0;	
Matches 1850;	Conservative	0; Mismatches	10; Indels 82; Gaps 1;
Qy	1	CCACGGCTCCGATGGCGACTCGGCCGCAACCTCGGCTTCAGAACCCGAGGCTGAGTCCA	60

```
QY 1059 TACAGTCAGTACCTAGATGACAACTGAAAGTAAAGGATTTGGTTTATCTGTACATAAC 1118
Db 1145 TACAGTCAGTACCTAGATGACAACTGAAAGTAAAGGATTTGGTTTATCTGTACATAAC 1204
QY 1119 CGCGCTATTAGAACTGCCCTGGACAGTTGCCCTTATGAGTCGGTACCTTGGCCATG 1178
Db 1205 CGCGCTATTAGAACTGCCCTGGACAGTTGCCCTTATGAGTCGGTACCTTGGCCATG 1264
QY 1179 GAGACATGAGCTTGCATCATCAAGTAATTTCTGTAACCTTCGAGAAAGCTTTGAATGCC 1238
Db 1265 GAGACATGAGCTTGCATCATCAAGTAATTTCTGTAACCTTCGAGAAAGCTTTGAATGCC 1324
QY 1239 GGCTTCATCCAGCCACTGATTATGATGGAGATTTGGCAGCATACCTTGATACCTGAGG 1298
Db 1325 GGCTTCATCCAGCCACTGATTATGATGGAGATTTGGCAGCATACCTTGATACCTGAGG 1384
QY 1299 AGAAGGTTGATTTCAACAAGACTCCAGTAAAGACTGGAGAGTTGAGGCGCCCTTT 1358
Db 1385 AGAAGGTTGATTTCAACAAGACTCCAGTAAAGACTGGAGAGTTGAGGCGCCCTTT 1444
QY 1359 ACTCGTCCCTTGGAGTATCTGAAGCAGAGAGTGGAAAGCGTTTCAATGAGAGTGGTGAT 1418
Db 1445 ACTCGTCCCTTGGAGTATCTGAAGCAGAGAGTGGAAAGCGTTTCAATGAGAGTGGTGAT 1504
QY 1419 CCAAGCTGCTGATTATGAGAACTCGGCTAGGATTTGAGCTCGACTGTGCAATAACATG 1478
Db 1505 CCAAGCTGCTGATTATGAGAACTCGGCTAGGATTTGAGCTCGACTGTGCAATAACATG 1564
QY 1479 CAGAAAGCTCGGAACCTCTGGGATAGCATCATCACCAGAGGAATGCCAAGTACGCCAC 1538
Db 1565 CAGAAAGCTCGGAACCTCTGGGATAGCATCATCACCAGAGGAATGCCAAGTACGCCAC 1624
QY 1539 ATGTGCTAGATTTACAACTGGAAGAGCTCATGTCACACCAGCACTGCCGGAAG 1598
Db 1625 ATGTGCTAGATTTACAACTGGAAGAGCTCATGTCACACCAGCACTGCCGGAAG 1684
QY 1599 GCTCTCACCGGGCGCTCAGTGCACCGAGTACTACCCAGAGCACGCTCGCGAAGTGTTA 1658
Db 1685 GCTCTCACCGGGCGCTCAGTGCACCGAGTACTACCCAGAGCACGCTCGCGAAGTGTTA 1744
QY 1659 CTCACCATGGAGAGGACAGAGGTTCTTTAGAGATTTGGATATAGCTGTCAGAAACT 1718
Db 1745 CTCACCATGGAGAGGACAGAGGTTCTTTAGAGATTTGGATATAGCTGTCAGAAACT 1804
QY 1719 GAAACCCGATTAGCTGCTCATATGACAGAGAAATGAAGGCTGCAGAGAAAGCAGCC 1778
Db 1805 GAAACCCGATTAGCTGCTCATATGACAGAGAAATGAAGGCTGCAGAGAAAGCAGCC 1864
QY 1779 CTTGTCAGCAGAAAGAAAGGCTGAACACGGAAAGAGCTCGGCTCGAGAGAAA 1838
Db 1865 CTTGTCAGCAGAAAGAAAGGCTGAACACGGAAAGAGCTCGGCTCGAGAGAAA 1924
QY 1839 GCGTTAAAAAAGAAAAAGA 1860
Db 1925 GCGTTAAAAAAGAAAAAGA 1946

RESULT 3
BC024279
LOCUS BC024279
DEFINITION Homo sapiens, 1938 bp mRNA linear HTC 28-FEB-2002
by T cells 3, clone IMAGE:4810833, mRNA.
ACCESSION BC024279
VERSION BC024279.1 GI:18999504
KEYWORDS HTC.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1938)
AUTHORS Strausberg, R.
TITLE Direct Submission
```

## JOURNAL

REMARK  
COMMENT

Submitted (26-FEB-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgabps-r@mail.nih.gov](mailto:cgabps-r@mail.nih.gov)  
Tissue Procurement: ATCC  
cDNA Library Preparation: Rubin Laboratory  
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Genome Sequence Centre, BC Cancer Agency, Vancouver, BC, Canada  
[info@bgsc.bc.ca](mailto:info@bgsc.bc.ca)  
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin, Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline Schein, Duane Smalius, Michael Smith, Lorraine Spence, Jeff Stott, Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAL Plate: 42 Row: 9 Column: 8  
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 7661951  
This clone has the following problem: frame shifted.

## FEATURES

## source

1..1938  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4810833"  
/tissue.type="placenta, choriocarcinoma"  
/clone.lib="NIH MGC\_21"  
/lab\_host="DH10B-R"  
/note="Vector: pOTB7"  
BASE COUNT 544 a 412 c 567 g 415 t  
ORIGIN

Query Match 46.1%; Score 1750.2; DB 11; Length 1938;  
Best Local Similarity 95.4%; Pred. No. 0;  
Matches 1847; Conservative 0; Mismatches 8; Indels 82; Gaps 1;  
QY 6 CGTCCGATGGCGACTCGCGCCGAAACCTCGGCTTCAGAACCCGAGGCTGAGTCCAAAGCT 65  
Db 1 CGCAAGATGGCGACTCGCGCCGAAACCTCGGCTTCAGAACCCGAGGCTGAGTCCAAAGCT 60  
QY 66 GGGCCCAAGGCTGACGGAGAGGAGGATGAGTTAAGCGGCTAGGACAGGAGAAAGGTG 125  
Db 61 GGGCCCAAGGCTGACGGAGAGGAGGATGAGTTAAGCGGCTAGGACAGGAGAAAGGTG 120  
QY 126 TTATCGCGGGCTGTGGCGCTCGACATACAGACCATGGGCGAGCTGGGATCAGCAG 185  
Db 121 TTATCGCGGGCTGTGGCGCTCGACATACAGACCATGGGCGAGCTGGGATCAGCAG 180  
QY 186 GAGGAAGCGCTGAGCGAGAGCGATGGGATGAGTACGCCATGGCTTCCTCCCGGAGAGC 245  
Db 181 GAGGAAGCGCTGAGCGAGAGCGATGGGATGAGTACGCCATGGCTTCCTCCCGGAGAGC 240  
QY 246 TCCCGCGGGAGTACAGTGGGAATATGACGAAGGAGGAGAGAGAGAGAGAGATTT 305  
Db 241 TCCCGCGGGAGTACAGTGGGAATATGACGAAGGAGGAGAGAGAGAGAGAGATTT 300  
QY 306 GAGAGACTGGAGGAGCAGTTGTCTATCAACGCTCTATGACTACAACTGCCATGTGGACTTG 365  
Db 301 GAGAGACTGGAGGAGCAGTTGTCTATCAACGCTCTATGACTACAACTGCCATGTGGACTTG 360  
QY 366 ATCAGACTGCTCAGGCTGGAAGGGAGCTTACCAAGGTGAGGATGCCCGCCAGAGATG 425  
Db 361 ATCAGACTGCTCAGGCTGGAAGGGAGCTTACCAAGGTGAGGATGCCCGCCAGAGATG 420  
QY 426 AGTGAATCTTTCCTTCTGACTGAAGAGCTCTGGCTGGAGTGGCTGCATCAGCAGATCAGC 485









Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 897)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone Distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM12220 row: m column: 06

High quality sequence stop: 744.

Location/Qualifiers

1. .897

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:5534381"

/clone\_lib="NIH\_MGC\_71"

/tissue\_type="leiomyosarcoma"

/lab\_host="DH10B (phage-resistant)"

/note="Organ: uterus; Vector: pCMV-SPORT6; Site\_1: NotI;

Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dt.

Average insert size 2.1 kb.

235 a 191 c 289 g 182 t

BASE COUNT  
ORIGIN

Query Match 21.3%; Score 810.6; DB 13; Length 897;

Best Local Similarity 96.3%; Pred. No. 3.4e-158;

Matches 852; Conservative 0; Mismatches 29; Indels 4; Gaps 2;

QY 6 CGTCCGATGGGACTCGCGCCCAACCTCGGTTTGAACCCGAGGCTAGTCCAGGCT 65

DB 1 CGCAAGATGGGACTCGCGCCCAACCTCGGTTTGAACCCGAGGCTAGTCCAGGCT 60

QY 66 GGGCCCAAGCTGTACGAGAGGAGGATGAGTTAAGCGGCTAGGACAAAGGAGAAGGTG 125

DB 61 GGGCCCAAGCTGTACGAGAGGAGGATGAGTTAAGCGGCTAGGACAAAGGAGAAGGTG 120

QY 126 TTATCGCGGGCTGTGGCCGCTCGACATACAAAGACCATGGGCCAGCGTGGATCAGCAG 185

DB 121 TTATCGCGGGCTGTGGCCGCTCGACATACAAAGACCATGGGCCAGCGTGGATCAGCAG 180

QY 186 GAGGAGGCGTGAGCGAGAGCGATGGGATGAGTACGCCATGGCTTCTCCCGGAGAGC 245

DB 181 GAGGAGGCGTGAGCGAGAGCGATGGGATGAGTACGCCATGGCTTCTCCCGGAGAGC 240

QY 246 TCCCGCGGGAGTACGAGTGGGAATATGACGAAGAGGAGGAGAAAACCCAGCTGGAGATT 305

DB 241 TCCCGCGGGAGTACGAGTGGGAATATGACGAAGAGGAGGAGAAAACCCAGCTGGAGATT 300

QY 306 GAGAGACTGGAGGAGCAGTTGTCTATCAACGCTCTATGACTACAACTGCCATGTGGACTTG 365

DB 301 GAGAGACTGGAGGAGCAGTTGTCTATCAACGCTCTATGACTACAACTGCCATGTGGACTTG 360

QY 366 ATCAGACTGCTCAGCTGGAAGGGAGCTTACCAGGTGAGGATGCCCGCCAGAGATG 425

DB 361 ATCAGACTGCTCAGCTGGAAGGGAGCTTACCAGGTGAGGATGCCCGCCAGAGATG 420

QY 426 ACTGAAATCTTTCCTTGAAGAGCTCTGGCTGGAGTGGCTGCATGACGAGATCAGC 485

DB 421 ACTGAAATCTTTCCTTGAAGAGCTCTGGCTGGAGTGGCTGCATGACGAGATCAGC 480

QY 486 ATGGCCCCAGGATGGCTGGACAGAGCAGCAGCTGTATGACCTCTTTGAGAAACCCGTTGAAG 545

DB 481 ATGGCCCCAGGATGGCTGGACAGAGCAGCAGCTGTATGACCTCTTTGAGAAACCCGTTGAAG 540

QY 546 GATTACATTTGCTTAACATTTGGCTAGAGTATGCCAGTACTACTCAGTTGGTGGGATTGGT 605

DB 546 GATTACATTTGCTTAACATTTGGCTAGAGTATGCCAGTACTACTCAGTTGGTGGGATTGGT 605

DB 541 GATTACATTTGCTTAACATTTGGCTAGAGTATGCCAGTACTCAGTTGGTGGGATTGGT 600

QY 606 CAGAAAGTGGGCTTGAGAAAGTTCCGCTCCGTTTGAAGGCTCTCTCGTCTGTGGT 665

DB 601 CAGAAAGTGGGCTTGAGAAAGTTCCGCTCCGTTTGAAGGCTCTCTCGTCTGTGGT 660

QY 666 TTACATATGACCAAAAGGACTCGCCCTCTCGGAGGCTTACCAGAGTTTGAAGTGGGATT 725

DB 661 TTACATATGACCAAAAGGACTCGCCCTCTCGGAGGCTTACCAGAGTTTGAAGTGGGATT 720

QY 726 GTGGAGTGTCTGGCTTGAGAAAGTCCACAGTCTTTCCGGGAGAGTTGGGATCCCA 785

DB 721 GTGGAGTGTCTGGCTTGAGAAAGTCCACAGTCTTTCCGGGAGAGTTGGGATCCCA 779

QY 786 CTCTATGATATGAGGCCACATTTGCAGAGTATGAAGATGTGACAGACCAATACCA 845

DB 780 CTCTATGATATGAGGCCACATTTGCAGAGTATGAAGATGTGACAGACCAATACCA 839

QY 846 GAG---TCAGTAATTCAGAACTATACAAAGCACTACACAGCTG 887

DB 840 CAGAGTCCAGTATTTTCAGAACTATACAAAGCACTACACAGCTG 884

RESULT 8

LOCUS BG682796

DEFINITION BG682796.1 GI:13914193

ACCESSION BG682796

VERSION BG682796.1

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

TITLE NIH-MGC http://mgc.nci.nih.gov/.

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone Distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLCMI615 row: p column: 22

High quality sequence stop: 856.

Location/Qualifiers

1. .861

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:4762245"

/clone\_lib="NIH\_MGC\_47"

/tissue\_type="neuroblastoma, cell line"

/lab\_host="DH10B (phage-resistant)"

/note="Organ: brain; Vector: pOTB7; Site\_1: XhoI; Site\_2:

EcoRI; cDNA made by oligo-dr priming. Directionally

cloned into EcoRI/XhoI sites using the following 5'

adaptor: GGCACGAG(G). Size-selected >500bp for average

insert size 1.8kb. Library constructed by Ling Hong in

the laboratory of Gerald M. Rubin (University of

California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies).

Note: this is a NIH\_MGC Library.

BASE COUNT 247 a 184 c 233 g 197 t

ORIGIN

Query Match 21.2%; Score 805.6; DB 12; Length 861;

Best Local Similarity 99.1%; Pred. No. 3.7e-157;

Matches 852; Conservative 0; Mismatches 4; Indels 4; Gaps 4;

```
QY 806 ATTTCCAGATATGAAGAATGTCAGAGACCCCAATACACAGTCAGTAAATTCAGAACTA 865
|||||
Db 2 ATTTCCAGATATGAAGAATGTCAGAGACCCCAATACACAGTCAGTAAATTCAGAACTA 61
|||||
QY 866 TAACAAGCACTACAGAGCTGGAGAAATATAAACCCCTATGAAGAAGCACTGTTGCCAGC 925
|||||
Db 62 TAACAAGCACTACAGAGCTGGAGAAATATAAACCCCTATGAAGAAGCACTGTTGCCAGC 121
|||||
QY 926 AGAGCACCAGGCTGCAGATATCAAGCATATATCGATTTTCAGATGAAATTCGCGA 985
|||||
Db 122 AGAGCACCAGGCTGCAGATATCAAGCATATATCGATTTTCAGATGAAATTCGCGA 181
|||||
QY 986 TCCTGCTCGCATTCAGTTGATCTTTGAGCGCGCCCTGGTCGAGAACTGCCCTGTCGCCAGA 1045
|||||
Db 182 TCCTGCTCGCATTCAGTTGATCTTTGAGCGCGCCCTGGTCGAGAACTGCCCTGTCGCCAGA 241
|||||
QY 1046 CTTATGGATCGTTACAGTCAGTACCTAGATCGACACTGAAAGTAAAGGATTTGGTTTT 1105
|||||
Db 242 CTTATGGATCGTTACAGTCAGTACCTAGATCGACACTGAAAGTAAAGGATTTGGTTTT 301
|||||
QY 1106 ATCTGTACATAACCGCTATTAGAACTGCCCTGGACAGTTCGCTTATGGAGTCGTA 1165
|||||
Db 302 ATCTGTACATAACCGCTATTAGAACTGCCCTGGACAGTTCGCTTATGGAGTCGTA 361
|||||
QY 1166 CCTCTTGCCCATGGAGAGACATGGAGTTGATCATCAAGTAATTTCTGTAACTTCGAGAA 1225
|||||
Db 362 CCTCTTGCCCATGGAGAGACATGGAGTTGATCATCAAGTAATTTCTGTAACTTCGAGAA 421
|||||
QY 1226 AGCTTTGATGCCGCTTCATCCAGGCCACTGATATGTGGAGATTTGGCAGGCATACCT 1285
|||||
Db 422 AGCTTTGATGCCGCTTCATCCAGGCCACTGATATGTGGAGATTTGGCAGGCATACCT 481
|||||
QY 1286 TGATTACCTGAGGAGAGGGTTGATTTCAACAAGACTCCAGTAAAGAGCTGGAGGAGTT 1345
|||||
Db 482 TGATTACCTGAGGAGAGGGTTGATTTCAACAAGACTCCAGTAAAGAGCTGGAGGAGTT 541
|||||
QY 1346 GAGGCGCGCTTTACTCGTGCTTTGGAGTATCTGAAGCAGGAGGTGGAAGAGCGTTTCAA 1405
|||||
Db 542 GAGGCGCGCTTTACTCGTGCTTTGGAGTATCTGAAGCAGGAGGTGGAAGAGCGTTTCAA 601
|||||
QY 1406 TGAGAGTGGTATCCAGCTCGGTGATATTCGAGAACTGGGCTAGGATTTAGGCTCGACT 1465
|||||
Db 602 TGAGAGTGGTATCCAGCTCGGTGATATTCGAGAACTGGGCTAGGATTTAGGCTCGACT 661
|||||
QY 1466 GTGCAATAACATGCAGAAAGCTCGGAACTCTGGATAGCATATGACCAGAGAAATGC 1525
|||||
Db 662 GTGCAATAACATGCAGAAAGCTCGGAACTCTGGATAGCATATGACCAGAGAAATGC 721
|||||
QY 1526 CAAGTACGCCAACATGTGGCTAGAGTATTACAACCT -GGAAAGAGCTCATGTTGACACCC 1584
|||||
Db 722 CAAGTACGCCAACATGTGGCTAGAGTATTACAAGCTGGGAAAGAGCTCATGTTGACACCC 781
|||||
QY 1585 A-GCACTCCCG -AAGGCTCTGACCGGGCGCTCGAGTGAC -AGTGACTACCCAGAGC 1641
|||||
Db 782 AGGCATCTGCCGGAAGGCTCTGACCGGGCGCTCGAGTGACCAAGGAGCTACCCAGAGC 841
|||||
QY 1642 ACGTCTGCGAAGTGTACTC 1661
|||||
Db 842 ACGTCTGCGAAGTGTACTC 861
|||||
```

```
RESULT 9
BM465186
LOCUS
DEFINITION AGNCOURT_6427767 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:5501938
5', mRNA sequence.
ACCESSION BM465186
VERSION BM465186.1 GI:18514228
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 989)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgabbs-r@mail.nih.gov](mailto:cgabbs-r@mail.nih.gov)  
Tissue Procurement: ATCC  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone Distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLAM12139 row: e column: 11  
High quality sequence start: 18  
High quality sequence stop: 718.  
Location/Qualifiers  
1. 989  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5501938"  
/clone\_lib="NIH\_MGC\_67"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: eye; Vector: pCMV-SPORT6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.75 kb. Library constructed by Life  
Technologies." 263 a 227 c 270 g 228 t 1 others

```
Query Match 21.2%; Score 805.2; DB 13; Length 989;  
Best Local Similarity 98.1%; Pred. No. 4.e-157;  
Matches 836; Conservative 0; Mismatches 13; Indels 3; Gaps 2;  
QY 718 GTGCGATTGTGGAAGCTGCTCGGCTTGAGAAGTCCACAGCTTTTCCGGCGACAGTTGG 777  
|||||  
Db 44 GTGCGATTGTGGAAGCTGCTCGGCTTGAGAAGTCCACAGCTTTTCCGGCGACAGTTGG 103  
|||||  
QY 778 CGATCCCACTCTATGATATGGAGGCCACATTTGCAGAGTATGAAGATGTCAGAGACC 837  
|||||  
Db 104 CGATCCCACTCTATGATATGGAGGCCACATTTGCAGAGTATGAAGATGTCAGAGACC 163  
|||||  
QY 838 CAATACCAAGCTCAGTAATTCAGAACTATAACAAGCACTACAGCAGCTGGAGAATATA 897  
|||||  
Db 164 CAATACCAAGCTCAGTAATTCAGAACTATAACAAGCACTACAGCAGCTGGAGAATATA 223  
|||||  
QY 898 AACCTATGAAGAAGCACTGTTGCAGGCGAGGCCAACAGGCTGGCAGAATATCAAGCAT 957  
|||||  
Db 224 AACCTATGAAGAAGCACTGTTGCAGGCGAGGCCAACAGGCTGGCAGAATATCAAGCAT 283  
|||||  
QY 958 ATATCGATTTTGAGATGAAATTTGGCATCTCTCGCATTCAGTTGATCTTTGAGCGCG 1017  
|||||  
Db 284 ATATCGATTTTGAGATGAAATTTGGCATCTCTCGCATTCAGTTGATCTTTGAGCGCG 343  
|||||  
QY 1018 CCCTGGTCCGAACTGCCCTTGTCCAGACTTATGGATCCGTACAGTCAGTACCTAGATC 1077  
|||||  
Db 344 CCCTGGTCCGAACTGCCCTTGTCCAGACTTATGGATCCGTACAGTCAGTACCTAGATC 403  
|||||  
QY 1078 GACAACTGAAAGTAAAGGATTTGGTTTATCTGTACATAAACCGGCTATTAGAAACTGCC 1137  
|||||  
Db 404 GACAACTGAAAGTAAAGGATTTGGTTTATCTGTACATAAACCGGCTATTAGAAACTGCC 463  
|||||  
QY 1138 CCTGGACAGTTGCCCTTATGGAGTCGGTACCTCTTGGCCCATGGAGAGATGGAGTTGATC 1197  
|||||  
Db 464 CCTGGACAGTTGCCCTTATGGAGTCGGTACCTCTTGGCCCATGGAGAGATGGAGTTGATC 523  
|||||  
QY 1198 ATCAAGTAATTTCTTAACCTTCGAGAAGCTTTGAATCGCGGCTTCATCCAGGCCACTG 1257  
|||||  
Db 524 ATCAAGTAATTTCTTAACCTTCGAGAAGCTTTGAATCGCGGCTTCATCCAGGCCACTG 583  
|||||  
QY 1258 ATTATGTGGAGATTTGGCAGGCATACCTTTGATTACCTGAGGAGAGGGTTGATTTCAAAC 1317  
|||||
```

```
|||||
584 ATTATGGAGATTGGCAGCATACCTGATTACCTGAGGAAGGGTTGATTTCAAC 643
|||||
1318 AAGACTCCAGTAAGAGCTGGAGGAGTTGAGGCGCGCTTTACTCGTGCTTGGAGTATC 1377
|||||
644 AAGACTCCAGTAAGAGCTGGAGGAGTTGAGGCGCGCTTTACTCGTGCTTGGAGTATC 703
|||||
1378 TGAAGCAGGAGGTGGAAGAGCGTTTCAATGAGAGTGGTGATCCAGCTGCGTATTATGC 1437
|||||
704 TGAAGCAGGAGGTGGGAGAGCGTTTCAATGAGAGTGGTGATCCAGCTGCGTATTATGC 763
|||||
1438 AGAAGTGGCTAGGATTGAGGCTCGACTGTGCAATACATGCAGAAAGCTCGGGAATCT 1497
|||||
764 AGAAGTGGCTAGGATTGAGGCTCGACTGTGCAATACATGCAGAAAGCTCGGGAATCT 823
|||||
1498 GGGATAGCATCATGACCA--GAGGAATGCCAAGTAC--GCCAATAGTGGCTAGAGTATT 1554
|||||
824 GGGATAGCATCATGACCAAGAGAAATGCCAGTACCGCAACATGGTGGCTAGAGTATT 883
|||||
1555 ACAAGCTGGAAA 1566
|||||
884 ACAAGCTGGAA 895

RESULT 10
AL563618/c 844 bp mRNA linear EST 16-FEB-2001
LOCUS AL563618 LTI_NFL001_NBC4 Homo sapiens cDNA clone CS0DD009YB08 3
DEFINITION prime, mRNA sequence.
ACCESSION AL563618
VERSION AL563618.1 GI:12913189
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 844)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 9106 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
source
1..844
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="LTI_NFL001_NBC4"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="PH10B"
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed
by Life Technologies. Contact : Feng Liang Life
Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
8371 Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 216 a 215 c 202 g 191 t 20 others
ORIGIN
Query Match 20.8%; Score 788.2; DB 9; Length 844;
Best Local Similarity 95.7%; Pred. No. 1.5e-153;
Matches 808; Conservative 19; Mismatches 15; Indels 2; Gaps 2;
2745 CAGCTCTCTACTGCTGCTGCTGCGCTGAGCGCCAGCGCTGAGCTCTCAGGCTGAG 2804
|||||
844 CAGCTCTCTACTGCTGCTGCTGCGCTGAGCGCCAGCGCTGAGCTCTCAGGCTGAG 785
|||||
```

QY	663	GGTTTACATATGACCAAGAGACTCGCCCTCTGGGAGGCTTACCGAGAGTTTG-AAAGTGC	721
Db	666	GGTTTACATATGACCAAGAGACTCGCCCTCTGGGAGGCTTACCGAGAGTTTGAAGTGC	725
QY	722	GATTGTGGAAGTGC-TCGGCTTGAGAAAGTCACAGTCTTTTCCGGCGACAGTTGGCGA	780
Db	726	CATTGTGGAAGTGCCTTCGGCTTGAGAAAGTNCACAGTCTTTTCCGGCGACAGTTGGCGA	785
QY	781	TCCCACTCTATGATGGAGGCGCCACATTTCCAGAGTATGAAGAAATGGTCAAGAAGACCCA	839
Db	786	TCCC-CTCTATGATATGGAGGNCCTTCACAGATATGAAAATGTCTANAGAGACCCA	843

[illegible]









GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 18, 2003, 13:45:33 ; Search time 44 Seconds  
(without alignments)  
2916.372 Million cell updates/sec

Title: US-09-763-985A-2

Perfect score: 4994

Sequence: 1 MATAAETSAEPEAEKAGP.....AATEAPKMSNADFAKFLRKLK 963

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_101002.\*

1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4994	100.0	963	21 AAY85422	Tumour antigen pro
2	4989	99.9	1179	22 AAE13123	Human SART3-IC-HER
3	2918	58.4	578	21 AAB56918	Human prostate can
4	526	10.5	104	22 ABB30113	Peptide #2764 enco
5	526	10.5	104	22 ABB35284	Peptide #2790 enco
6	526	10.5	104	22 ABB20728	Protein #2727 enco
7	526	10.5	104	22 AAM56114	Human brain expres
8	526	10.5	104	22 AAM68486	Human bone marrow
9	526	10.5	104	22 AAM16294	Peptide #2728 enco
10	526	10.5	104	22 AAM28783	Peptide #2820 enco

11	526	10.5	104	22 AAM04028	Peptide #2710 enco
12	526	10.5	104	23 ABG38067	Human peptide enco
13	513	10.3	108	23 ABP33657	Human ORE2630 prot
14	425	8.5	941	22 ABB59801	Drosophila melanog
15	420	8.4	100	21 AAG01853	Human secreted pro
16	272	5.4	414	17 AAM10529	Saccharomyces cere
17	272	5.4	414	22 AAB30809	Amino acid sequenc
18	272	5.4	1305	22 ABB67385	Drosophila melanog
19	268.5	5.4	702	22 ABB59602	Drosophila melanog
20	259.5	5.2	691	21 AAG48233	Arabidopsis thalia
21	259.5	5.2	705	21 AAG48232	Arabidopsis thalia
22	245.5	4.9	536	22 AAB92836	Human protein sequ
23	245.5	4.9	707	16 AAR79912	Human nucleolin.
24	245.5	4.9	707	20 AAW84052	Human V3 loop HIV
25	245.5	4.9	707	22 AAB48964	Human nucleolin.
26	241.5	4.8	1009	22 ABB58338	Drosophila melanog
27	236.5	4.7	673	21 AAG39359	Arabidopsis thalia
28	236.5	4.7	683	21 AAG39358	Arabidopsis thalia
29	236	4.7	569	21 AAG48234	Arabidopsis thalia
30	234	4.7	836	22 AAM40367	Human polypeptide
31	229.5	4.6	582	21 AAG39360	Arabidopsis thalia
32	229.5	4.6	650	21 AAG52494	Arabidopsis thalia
33	229.5	4.6	662	21 AAG52493	Arabidopsis thalia
34	229.5	4.6	675	21 AAG52492	Arabidopsis thalia
35	225	4.5	708	21 AAB43567	Human cancer assoc
36	224	4.5	483	14 AAR43893	elav. Drosophila
37	219.5	4.4	721	22 ARG15463	Novel human diagno
38	217.5	4.4	1972	17 AAM00024	Smooth muscle myos
39	217	4.3	41	22 ABB28701	Peptide #1352 enco
40	217	4.3	41	22 ABB33883	Peptide #1389 enco
41	217	4.3	41	22 ABB19324	Protein #1323 enco
42	217	4.3	41	22 AAM54649	Human brain expres
43	217	4.3	41	22 AAM67053	Human bone marrow
44	217	4.3	41	22 AAM14915	Peptide #1349 enco
45	217	4.3	41	22 AAM27344	Peptide #1381 enco

#### ALIGNMENTS

##### RESULT 1

AAY85422

ID AAY85422 standard; Protein; 963 AA.

XX AC AAY85422;

XX DT 23-JUN-2000 (first entry)

XX DE Tumour antigen protein SART-3.

XX DE Tumour antigen protein SART-3.

XX KW Tumour antigen protein; SART-3; cytotoxic T-cell; HLA antigen; tumour.

XX OS Homo sapiens.

XX WO200012701-A1.

XX PD 09-MAR-2000.

XX PF 27-AUG-1999; 99WO-JP04622.

XX PR 28-AUG-1998; 98JP-0242660.

XX PA (SUMI ) SUMITOMO PHARM CO LTD.

XX PA (ITOH/) ITOH K.

XX PI Itoh K, Nakao M;

XX DR WPI; 2000-237868/20.

XX DR N-PSDB; AA290783.

XX PT Tumor antigen protein SART-3 recognized by cytotoxic T-cells binding to  
XX HLA antigen for treatment and diagnosis of tumors  
XX

PS Claim 1; Page 54-58; 89pp; Japanese.

XX This represents a tumour antigen protein (SART-3) which is recognized by  
CC cytotoxic T-cells binding to HLA antigen. SART-3 can be expressed by  
CC standard recombinant methodology. SART-3 and its peptide derivatives can  
CC be used in the prevention, treatment and diagnosis of tumours in vitro or  
CC in vivo.

XX  
SQ Sequence 963 AA;

Query Match 100.0%; Score 4994; DB 21; Length 963;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 963; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MATAETSASEPEAESKAGPKADGEDEVKAARTRRKVLRSRAVAATAATYKTMGPAMDQEE 60  
DB 1 MATAETSASEPEAESKAGPKADGEDEVKAARTRRKVLRSRAVAATAATYKTMGPAMDQEE 60

QY 61 GVSDEGDEYAMASSAESPGYEWYDEEEKNQLEIERLEEQLSINVDYNCHVDLIR 120  
DB 61 GVSDEGDEYAMASSAESPGYEWYDEEEKNQLEIERLEEQLSINVDYNCHVDLIR 120

QY 121 LRLLEGELTKVRMAROKMSEIPLTEELMLELHDEISMAQDGLDREHYVDLFEKAVDY 180  
DB 121 LRLLEGELTKVRMAROKMSEIPLTEELMLELHDEISMAQDGLDREHYVDLFEKAVDY 180

QY 181 ICPNIWLEXGOYSGVGIGKGGLEKVRVSFERALSSVGLHMTKGLALWEAYREFESAIVE 240  
DB 181 ICPNIWLEXGOYSGVGIGKGGLEKVRVSFERALSSVGLHMTKGLALWEAYREFESAIVE 240

QY 241 AARLEKVLHFRRLQALPIYDMEATPAEYEESEDDIPESVTONYNKALQOLEKYPYEE 300  
DB 241 AARLEKVLHFRRLQALPIYDMEATPAEYEESEDDIPESVTONYNKALQOLEKYPYEE 300

QY 301 ALLOAEAPRLAEYOVIDFEMKIGDPARLQIFERALVENCIVLPDLWIRYSQYLDRLQKV 360  
DB 301 ALLOAEAPRLAEYOVIDFEMKIGDPARLQIFERALVENCIVLPDLWIRYSQYLDRLQKV 360

QY 361 KDLVLSVHNRAIRNCPTWALMSRYLLAMERHGVHDSVTFERKALNAGFTQATDYVEI 420  
DB 361 KDLVLSVHNRAIRNCPTWALMSRYLLAMERHGVHDSVTFERKALNAGFTQATDYVEI 420

QY 421 WQAYLDLRRRVDFKODSKELEELRAAFTRALEYLKQVEERFNEGSDPSCVIMONWAR 480  
DB 421 WQAYLDLRRRVDFKODSKELEELRAAFTRALEYLKQVEERFNEGSDPSCVIMONWAR 480

QY 481 IEARLCNNMOKARELWDSITWTRGNKAYANWLEYNLERAHGDTQCHRALHRAVQCTSD 540  
DB 481 IEARLCNNMOKARELWDSITWTRGNKAYANWLEYNLERAHGDTQCHRALHRAVQCTSD 540

QY 541 YPEHVCEVLLTMEETEGSLEDWDIAVQKTETRLARVNEQRMKAKEAALVQOEEKABQ 600  
DB 541 YPEHVCEVLLTMEETEGSLEDWDIAVQKTETRLARVNEQRMKAKEAALVQOEEKABQ 600

QY 601 RRRARAEKALKKKKIRGPERKGADEDEKQWGDDEEQPSKRRRVENSIPAAGETQNY 660  
DB 601 RRRARAEKALKKKKIRGPERKGADEDEKQWGDDEEQPSKRRRVENSIPAAGETQNY 660

QY 661 EVAAGPAGCAADVVEPPSKQKEAASLKRDMPKVLHDSKDSITVFSNLPYSNOEPT 720  
DB 661 EVAAGPAGCAADVVEPPSKQKEAASLKRDMPKVLHDSKDSITVFSNLPYSNOEPT 720

QY 721 KLRLFEACEGVVQIRPISNDRGDFRGYCYVEFEKESKALQALEMDRKSVEGRPMFVSPC 780  
DB 721 KLRLFEACEGVVQIRPISNDRGDFRGYCYVEFEKESKALQALEMDRKSVEGRPMFVSPC 780

QY 781 VDKSNPDPKVPYRSTSLKHKLFISGLPFSCTKEELEICAHGTVKDLRLVTNRAGKP 840  
DB 781 VDKSNPDPKVPYRSTSLKHKLFISGLPFSCTKEELEICAHGTVKDLRLVTNRAGKP 840

QY 841 KGLAVVEYENESQASQAVNMKGDMTKENIKVAISNPPORVPEKPTRKAPGPMLLP 900  
DB 841 KGLAVVEYENESQASQAVNMKGDMTKENIKVAISNPPORVPEKPTRKAPGPMLLP 900

QY 901 QTYGARGKGTQLSLPRALQRPSSAAAPQAENGPAAPAAVAPAAATEAPKMSNADFAKLF 960  
DB 901 QTYGARGKGTQLSLPRALQRPSSAAAPQAENGPAAPAAVAPAAATEAPKMSNADFAKLF 960

QY 961 LRK 963  
DB 961 LRK 963

RESULT 2  
AAE13123  
ID AAE13123 standard; Protein; 1179 AA.

XX AC AAE13123;  
XX AC AAE13123;  
XX 28-JAN-2002 (first entry)  
XX Human SART3-IC-HER-2 membrane distal intracellular domain fusion protein.  
XX Immunostimulatory fusion protein; IFP; antigen component; therapy;  
XX immunostimulatory component; T-cell mediated immune response; DC;  
XX dendritic cell; colon cancer; breast carcinoma; ovarian cancer;  
XX human; HER-2 membrane distal intracellular domain; SART3-IC;  
XX squamous cell carcinoma antigen; fusion protein.  
XX Homo sapiens.  
XX WO200174855-A2.  
XX 11-OCT-2001.  
XX 30-MAR-2001; 2001WO-US10515.  
XX 30-MAR-2000; 2000US-193504P.  
XX (DEND-) DENDREON CORP.  
XX Laus R, Vidovic D, Graddis T;  
XX WPI: 2001-662965/76.  
XX N-PSDB; AAD21574.

XX An immunostimulatory fusion protein comprising the intracellular domain  
XX of HER-2 and an antigen elicits an immune response to the antigen and  
XX is useful for the treatment of associated cancer associated -  
XX Disclosure; Page 55-58; 59pp; English.

XX The invention relates to immunostimulatory fusion proteins (IFP) and  
XX nucleic acid molecules encoding such proteins. The IFPs comprise a  
XX polypeptide antigen component and an immunostimulatory component derived  
XX from the intracellular domain of HER-2 protein which is effective to  
XX elicit a protective dendritic cell (DC)-induced T-cell mediated cellular  
XX immune response to the antigen. IFP or superactivated dendritic cells  
XX are used to treat cancer e.g. breast carcinoma, ovarian and colon cancer  
XX associated with a particularly antigen. The present sequence is a  
XX fusion protein which comprises human squamous cell  
XX carcinoma antigen, SART3-IC and mature human HER-2 membrane distal  
XX intracellular domain.

SQ Sequence 1179 AA;

Query Match 99.9%; Score 4989; DB 22; Length 1179;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 962; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MATAETSASEPEAESKAGPKADGEDEVKAARTRRKVLRSRAVAATAATYKTMGPAMDQEE 60  
DB 1 MATAETSASEPEAESKAGPKADGEDEVKAARTRRKVLRSRAVAATAATYKTMGPAMDQEE 60

QY 61 GVSDEGDEYAMASSAESPGYEWYDEEEKNQLEIERLEEQLSINVDYNCHVDLIR 120  
DB 61 GVSDEGDEYAMASSAESPGYEWYDEEEKNQLEIERLEEQLSINVDYNCHVDLIR 120

Db 61 GVSDEGDEYAMASSAESPGYEWYDEEKEKQLEIERLEEQLSINVYDYNCHVDLIR 120  
QY 121 LRLREGELTKVMARQKMSSEIFPLTEELWLEHLDEISMAODGLDREHVDLFEKAVXDY 180  
Db 121 LRLREGELTKVMARQKMSSEIFPLTEELWLEHLDEISMAODGLDREHVDLFEKAVXDY 180  
QY 181 ICPNTWLEYGOYVSGGIGOGKLEKRVSVFERRALSSVGLHMTKGLALWEAYREFESAIVE 240  
Db 181 ICPNTWLEYGOYVSGGIGOGKLEKRVSVFERRALSSVGLHMTKGLALWEAYREFESAIVE 240  
QY 241 AARLEKVSHPRLQALPILYDMEATFAEYEEHSEDPPIPESVIQTNYNKALQLEKYPYEE 300  
Db 241 AARLEKVSHPRLQALPILYDMEATFAEYEEHSEDPPIPESVIQTNYNKALQLEKYPYEE 300  
QY 301 ALLOAEAPRLAEYQAYIDFEMKIGDPARIQLIFERALVENCILVPLDIWIRYSOYLDRLQKV 360  
Db 301 ALLOAEAPRLAEYQAYIDFEMKIGDPARIQLIFERALVENCILVPLDIWIRYSOYLDRLQKV 360  
QY 361 KDLVLSVHNRAIRNCPWTVALWSRYLLAMERHGVHDQVTSVTFKALNAGFTQATDYVEI 420  
Db 361 KDLVLSVHNRAIRNCPWTVALWSRYLLAMERHGVHDQVTSVTFKALNAGFTQATDYVEI 420  
QY 421 WQAYLDYLRRRVDFKODSSKELEELRAAFTRALEYLKOEVERFNESEDPSCVIMQNWAR 480  
Db 421 WQAYLDYLRRRVDFKODSSKELEELRAAFTRALEYLKOEVERFNESEDPSCVIMQNWAR 480  
QY 481 IEARLCNNMOKARELWDSIMTRGNKAYANMWLEYNLERAHGDTQHCRAKALHRAVQCTSD 540  
Db 481 IEARLCNNMOKARELWDSIMTRGNKAYANMWLEYNLERAHGDTQHCRAKALHRAVQCTSD 540  
QY 541 YPEHVEVLLTWTEGSLDWDIAVQKTETRLARVNEQRMKAKEAALVQOEEKEAQ 600  
Db 541 YPEHVEVLLTWTEGSLDWDIAVQKTETRLARVNEQRMKAKEAALVQOEEKEAQ 600  
QY 601 RKRAREAKKALKKKKIRGPEKRGADDEDEKWDDEEOPSKRRRVENSIPAAGETQNV 660  
Db 601 RKRAREAKKALKKKKIRGPEKRGADDEDEKWDDEEOPSKRRRVENSIPAAGETQNV 660  
QY 661 EVAAGPAGCAADVPEPPSKOKAKAASLKRDMPKVLDHSSKDSITVFFVSNLPSYMOEPT 720  
Db 661 EVAAGPAGCAADVPEPPSKOKAKAASLKRDMPKVLDHSSKDSITVFFVSNLPSYMOEPT 720  
QY 721 KLRPLFEAGGEVVOIRPISNRGDFRGYCYVEFEKESALQALEMDRKSVEGRPMFVSPC 780  
Db 721 KLRPLFEAGGEVVOIRPISNRGDFRGYCYVEFEKESALQALEMDRKSVEGRPMFVSPC 780  
QY 781 VDKSKNPDKVPRYSTSLEKHKLFISGLPFSCTKEELEICAKHGTVDLRLVTNRAGKP 840  
Db 781 VDKSKNPDKVPRYSTSLEKHKLFISGLPFSCTKEELEICAKHGTVDLRLVTNRAGKP 840  
QY 841 KGLAYVEVENESQAQVAMKMDGMTIKENIKVAISNPPORVPEKPEKTRKAPGPMLLP 900  
Db 841 KGLAYVEVENESQAQVAMKMDGMTIKENIKVAISNPPORVPEKPEKTRKAPGPMLLP 900  
QY 901 QTYGARGKRTOLSLPRALORPSAAAPQAEANGPAAAPAAVAPAAATEAPKMSNADFAKLF 960  
Db 901 QTYGARGKRTOLSLPRALORPSAAAPQAEANGPAAAPAAVAPAAATEAPKMSNADFAKLF 960  
QY 961 LR 962  
Db 961 LR 962

RESULT 3

AAB56918

ID AAB56918 standard; Protein; 578 AA.

XX AC AAB56918;

XX AC 13-MAR-2001 (first entry)

DT DT Human prostate cancer antigen protein sequence SEQ ID NO:1496.

XX DE

Human; prostate cancer; prostate cancer antigen; detection; diagnosis; neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular; vulnary; gastrointestinal; nephrotropic; antiinfective; gynaecological; antibacterial; gene therapy; neural; immune; reproductive; renal; gastrointestinal; pulmonary; cardiovascular; proliferative disorder; wound; infectious disease.  
OS Homo sapiens.  
PN WO200055174-A1.  
XX 21-SEP-2000.  
XX 08-MAR-2000; 2000WO-US05988.  
XX 12-MAR-1999; 99US-0124270.  
PR (HUMA-) HUMAN GENOME SCI INC.  
PA (ROSE/) ROSEN C A.  
XX Rosen CA, Ruben SM;  
PI WPI; 2000-587513/55.  
DR N-PSDB; AAF16121.  
XX Prostate cancer associated gene sequences, referred to as prostate cancer antigens, useful for treatment, prevention, and diagnosis of disorders such as prostate cancer -  
Claim 11; Page 1934-1936; 2338pp; English.  
XX AAF15566 to AAF16505 encode the human prostate cancer associated proteins, called prostate cancer antigens, given in AAB56363 to AAB57302. The prostate cancer antigens can have neuroprotective, cytostatic, cardioactive, immunomodulatory, muscular, vulnary, gastrointestinal, nephrotropic, antiinfective, gynaecological and antibacterial activities, and can be used in gene therapy. The prostate cancer antigen polynucleotides may be used for detection of prostate cancer, chromosome identification, as chromosome markers, and for numerous other diagnostic or research purposes. The prostate cancer antigens may be used to treat disorders such as neural, immune, muscular, reproductive, gastrointestinal, pulmonary, cardiovascular, renal, and proliferative disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to AAB57303 represent sequences used in the exemplification of the present invention.  
SQ Sequence 578 AA;  
Query Match 58.4%; Score 2918; DB 21; Length 578;  
Best Local Similarity 99.1%; Pred. No. 5e-223;  
Matches 564; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 395 DHQVTSVTFEKAALNAGFIQATDYVEIWOAYLDYLRVRVDFKODSSKELEELRAAFTRALE 454  
Db 10 DFSPTVPTFEKAALNAGFIQATDYVEIWOAYLDYLRVRVDFKODSSKELEELRAAFTRALE 69  
QY 455 YLKQVEEERFNESEDPSCVIMONWARIARLNNMOKARELWDSIMTRNNAKYANMWLEY 514  
Db 70 YLKQVEEERFNESEDPSCVIMONWARIARLNNMOKARELWDSIMTRNNAKYANMWLEY 129  
QY 515 YNLERAHGDTQHCRAKALHRAVQCTSDYPEHVCVEVLLTMBTEGSLDWDIAVQKTETRLA 574  
Db 130 YNLERAHGDTQHCRAKALHRAVQCTSDYPEHVCVEVLLTMBTEGSLDWDIAVQKTETRLA 189  
QY 575 RVNEQRMKAKEAALVQOEEKEAQKRAEKKALKKKKIRGPEKRGADDEDEKENG 634  
Db 190 RVNEQRMKAKEAALVQOEEKEAQKRAEKKALKKKKIRGPEKRGADDEDEKENG 249  
QY 635 DDEEOPSKRRRVENSIPAAGETQNVVEAAGPAGCAADVPEPPSKOKAKAASLKRDMPK 694  
Db 250 DDEEOPSKRRRVENSIPAAGETQNVVEAAGPAGCAADVPEPPSKOKAKAASLKRDMPK 309  
QY 695 VLHDSSKDSITVFFVSNLPSYMOEPTDKLRPLFEACGEVVOIRPISNRGDFRGYCYVEFK 754





RESULT 8  
 AAM68486  
 ID AAM68486 standard; Protein; 104 AA.  
 XX  
 AC AAM68486;  
 XX  
 DT 06-NOV-2001 (first entry)  
 XX  
 DE Human bone marrow expressed probe encoded protein SEQ ID NO: 28792.  
 XX  
 KW Human; bone marrow expressed exon; gene expression analysis; probe;  
 KW microarray; cancer; leukaemia; lymphoma; myeloma.  
 XX  
 OS Homo sapiens.  
 PN WO200157276-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 30-JAN-2001; 2001WO-US00668.  
 XX  
 PR 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 DR WPI; 2001-488900/53.  
 XX  
 PT Human genome-derived single exon nucleic acid probes useful for  
 PT analyzing gene expression in human bone marrow -  
 XX  
 PS Example 4; SEQ ID NO: 28792; 658pp + Sequence Listing; English.  
 XX  
 CC The present invention provides a number of single exon nucleic acid  
 CC probes which are derived from genomic sequences expressed in the human  
 CC bone marrow. They can be used to measure gene expression in bone marrow  
 CC samples, which may enable the improved diagnosis and treatment of cancers  
 CC such as lymphoma, leukaemia and myeloma. The present sequence is a  
 CC protein encoded by one of the probes of the invention.  
 XX  
 SQ Sequence 104 AA;  
 Query Match 10.5%; Score 526; DB 22; Length 104;  
 Best Local Similarity 100.0%; Pred. No. 5.2e-34;  
 Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MATAAETSASEPEAESKAGPKADGEEDEVKAARTRRKVLRAVAATAATYKTMGPAMDQEE 60  
 DB 1 MATAAETSASEPEAESKAGPKADGEEDEVKAARTRRKVLRAVAATAATYKTMGPAMDQEE 60  
 QY 61 GVSESDGDEYAMASSAESSPGEYEWYDEEEKNLEIERLEEQ 104  
 DB 61 GVSESDGDEYAMASSAESSPGEYEWYDEEEKNLEIERLEEQ 104  
 RESULT 9  
 AAM16294  
 ID AAM16294 standard; Protein; 104 AA.  
 XX  
 AC AAM16294;  
 XX  
 DT 12-OCT-2001 (first entry)  
 XX  
 DE Peptide #2728 encoded by probe for measuring cervical gene expression.  
 XX

KW Probe; human; microarray; gene expression; cervical epithelial cell;  
 KW cervical cancer.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157278-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 30-JAN-2001; 2001WO-US00670.  
 XX  
 PR 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 DR WPI; 2001-488901/53.  
 XX  
 PT Human genome-derived single exon nucleic acid probes useful for  
 PT analyzing gene expression in human cervical epithelial cells -  
 XX  
 PS Claim 27; SEQ ID No 21120; 487pp; English.  
 XX  
 CC The present invention relates to human single exon nucleic acid probes  
 CC (SENP: see AAI10068-AAI28459). The present sequence is a peptide encoded  
 CC by one such probe. The SENPs are derived from human Hela cells. The SENPs  
 CC can be used to produce a single exon microarray, which can be used for  
 CC measuring human gene expression in a sample derived from human cervical  
 CC epithelial cells. By measuring gene expression, the probes are therefore  
 CC useful in grading and/or staging of diseases of the cervix, notably  
 CC cervical cancer.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 104 AA;  
 Query Match 10.5%; Score 526; DB 22; Length 104;  
 Best Local Similarity 100.0%; Pred. No. 5.2e-34;  
 Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MATAAETSASEPEAESKAGPKADGEEDEVKAARTRRKVLRAVAATAATYKTMGPAMDQEE 60  
 DB 1 MATAAETSASEPEAESKAGPKADGEEDEVKAARTRRKVLRAVAATAATYKTMGPAMDQEE 60  
 QY 61 GVSESDGDEYAMASSAESSPGEYEWYDEEEKNLEIERLEEQ 104  
 DB 61 GVSESDGDEYAMASSAESSPGEYEWYDEEEKNLEIERLEEQ 104  
 RESULT 10  
 AAM28783  
 ID AAM28783 standard; Protein; 104 AA.  
 XX  
 AC AAM28783;  
 XX  
 DT 17-OCT-2001 (first entry)  
 XX  
 DE Peptide #2820 encoded by probe for measuring placental gene expression.  
 XX  
 KW Probe; microarray; human; placenta; antenatal diagnosis;  
 KW genetic disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157272-A2.

```

XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00663.
XX PF 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488897/53.
XX PT Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human placenta.
XX PS Claim 27; SEQ ID No 29052; 654pp; English.
XX CC The present invention relates to single exon nucleic acid probes (SENP:
XX CC see AA131315-AA157546). The present sequence is a peptide encoded by one
XX CC such probe. The probes are useful for producing a microarray for
XX CC predicting, measuring and displaying gene expression in samples derived
XX CC from human placenta. The probes are useful for antenatal diagnosis of
XX CC human genetic disorders.
XX SQ Sequence 104 AA;
XX Query Match 10.5%; Score 526; DB 22; Length 104;
XX Best Local Similarity 100.0%; Pred. No. 5.2e-34;
XX Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MATAAETSASEPEAESKAGPKADGEDEVKAAARTRRKVLRAVAATAATYKTMGPAWDQOEE 60
DB 1 MATAAETSASEPEAESKAGPKADGEDEVKAAARTRRKVLRAVAATAATYKTMGPAWDQOEE 60
QY 61 GYSESDGDEYAMASSAESPGYEWYDEEEKNQLEIRLEE 104
DB 61 GYSESDGDEYAMASSAESPGYEWYDEEEKNQLEIRLEE 104
RESULT 11
AAM04028
ID AAM04028 standard; Protein; 104 AA.
XX AC AAM04028;
XX DT 09-OCT-2001 (first entry)
XX DE Peptide #2710 encoded by probe for measuring breast gene expression.
XX KW Probe; human; breast disease; breast cancer; development disorder;
XX KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX OS Homo sapiens.
XX PN WO200157270-A2.
XX PD 09-AUG-2001.
XX PF 29-JAN-2001; 2001WO-US00661.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.

```

```

PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-476286/51.
XX PT Novel single exon nucleic acid probe used to measuring gene expression
XX PT in a human breast.
XX PS Claim 27; SEQ ID No 12768; 322pp; English.
XX CC The present invention relates to novel single exon nucleic acid probes
XX CC (see AA100010-AA110067). The present sequence is a peptide encoded by one
XX CC such probe. The probes are useful for measuring human gene expression in
XX CC a human breast sample, where the probe hybridises at high stringency to a
XX CC nucleic acid expressed in the human breast. The probes are useful for
XX CC predicting, diagnosing, grading, staging, monitoring and prognosing
XX CC diseases of the human breast, particularly those diseases with polygenic
XX CC aetiology. The diseases include: breast cancer, disorders of development,
XX CC inflammatory diseases of the breast, fibrocystic changes, proliferative
XX CC breast disease and non-carcinoma tumours.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 104 AA;
XX Query Match 10.5%; Score 526; DB 22; Length 104;
XX Best Local Similarity 100.0%; Pred. No. 5.2e-34;
XX Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MATAAETSASEPEAESKAGPKADGEDEVKAAARTRRKVLRAVAATAATYKTMGPAWDQOEE 60
DB 1 MATAAETSASEPEAESKAGPKADGEDEVKAAARTRRKVLRAVAATAATYKTMGPAWDQOEE 60
QY 61 GYSESDGDEYAMASSAESPGYEWYDEEEKNQLEIRLEE 104
DB 61 GYSESDGDEYAMASSAESPGYEWYDEEEKNQLEIRLEE 104
RESULT 12
ABG38067
ID ABG38067 standard; Peptide; 104 AA.
XX AC ABG38067;
XX DT 19-AUG-2002 (first entry)
XX DE Human peptide encoded by genome-derived single exon probe SEQ ID 27732.
XX KW Human; single exon probe; asthma; lung cancer; COPD; ILD;
XX KW chronic obstructive pulmonary disease; interstitial lung disease;
XX KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
XX KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
XX KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
XX KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;
XX KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
XX KW primary ciliary dyskinesia; pulmonary hypertension;
XX KW hyaline membrane disease.
XX OS Homo sapiens.
XX PN WO200186003-A2.
XX PD 15-NOV-2001.
XX PF 30-JAN-2001; 2001WO-US00665.
XX PR 04-FEB-2000; 2000US-180312P.
XX PR 26-MAY-2000; 2000US-207456P.
XX PR 30-JUN-2000; 2000US-0608408.

```

PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-234687P.  
 PR 27-SEP-2000; 2000US-236359P.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX (MOLE-) MOLECULAR DYNAMICS INC.  
 PA Penn SG, Hanzel DK, Chen W, Rank DR;  
 PI WPI; 2002-114183/15.  
 XX  
 DR Spatially-addressable set of single exon nucleic acid probes, used to  
 XX measure gene expression in human lung samples -  
 XX  
 PT Claim 27; SEQ ID NO 27732; 634pp; English.  
 PS  
 XX The invention relates to a spatially-addressable set of single exon  
 CC nucleic acid probes for measuring gene expression in a sample derived  
 CC from human lung comprising single exon nucleic acid probes having one of  
 CC 12614 nucleic acid sequences mentioned in the specification, or their  
 CC complements or the 12387 open reading frames derived from the 12614  
 CC probes. Also included are a microarray comprising the novel set of  
 CC probes; the novel set of probes which hybridise at high stringency to a  
 CC nucleic acid expressed in the human lung; measuring gene expression in a  
 CC sample derived from human lung, comprising (a) contacting the array with  
 CC a collection of detectably labeled nucleic acids derived from human lung  
 CC mRNA, and (b) measuring the label detectably bound to each probe of  
 CC the array; identifying exons in a eukaryotic genome, comprising  
 CC (a) algorithmically predicting at least one exon from genomic sequences  
 CC of the eukaryote; and (b) detecting specific hybridisation of detectably  
 CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,  
 CC having a fragment identical to the predicted exon, the probe is included  
 CC in the above mentioned microarray; assigning exons to a single gene,  
 CC comprising (a) identifying exons from genomic sequence by the method  
 CC above and (b) measuring the expression of each of the exons in several  
 CC tissues and/or cell types using hybridisation to a single exon  
 CC microarrays having a probe with the exon, where a common pattern of  
 CC expression of the exons in the tissues and/or cell types indicates that  
 CC the exons should be assigned to a single gene; a peptide comprising one  
 CC of 12011 sequences, mentioned in the specification, or encoded by the  
 CC probes/open reading frames (ORF). The probes are used for gene  
 CC expression analysis, and for identifying exons in a gene, particularly  
 CC using human lung derived mRNA and for the study of lung diseases  
 CC such as asthma, lung cancer, chronic obstructive pulmonary disease  
 CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary  
 CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,  
 CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary  
 CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,  
 CC pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic  
 CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension  
 CC and hyaline membrane disease. The present sequence is a peptide/protein  
 CC encoded by a single exon probe of the invention.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 104 AA;  
 Query Match 10.5%; Score 526; DB 23; Length 104;  
 Best Local Similarity 100.0%; Pred. No. 5.2e-34;  
 Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MATAAETASAEPAESKAGPKADGEDEYKAAATRRKVLRSRAVAATYKTMGPANDQEE 60  
 DQ 1 MATAAETASAEPAESKAGPKADGEDEYKAAATRRKVLRSRAVAATYKTMGPANDQEE 60  
 QY 61 GVSSEGDGYAMASSAESPGYEWYDEEEKNLEIRLEEQ 104  
 DB 61 GVSSEGDGYAMASSAESPGYEWYDEEEKNLEIRLEEQ 104

RESULT 13

b. . . . .

ABP33657  
 ID ABP33657 standard; Protein; 108 AA.  
 XX  
 AC ABP33657;  
 XX  
 DT 08-JUL-2002 (first entry)  
 DE  
 DE Human ORF2630 protein, SEQ ID NO:5260.  
 XX  
 KW Human: ORF; open reading frame; ORFX; drug screening; diagnosis;  
 KW disease monitoring; cytokine; cell proliferation; cell differentiation;  
 KW immune modulation; haematopoiesis regulation; tissue growth;  
 KW angiogenesis; activin; inhibin; chemotactic; chemokinetic; haemostatic;  
 KW thrombolytic; tumour inhibition; bodily characteristic; fertility;  
 KW behaviour; cancer; proliferative disorder; neurological disorder;  
 KW cardiovascular disease; immune system disorder; organ transplantation;  
 KW tissue growth disorder; tissue regeneration disorder; diabetes mellitus;  
 KW hypothyroidism; cholesterol ester storage disease; infection; vulnery;  
 KW vasotropic; antipsoriatic; antidiabetic; cytostatic; neurotropic;  
 KW neuroprotective; antiatherosclerotic; anticoagulant; thrombolytic;  
 KW cardiant; hypotensive; antithyroid; antiinflammatory; immunomodulator;  
 KW dermatological; analgesic; virucide; antibacterial; fungicide.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200190366-A2.  
 XX  
 PD 29-NOV-2001.  
 XX  
 PF 24-MAY-2001; 2001WO-US17076.  
 XX  
 PR 24-MAY-2000; 2000US-206690P.  
 XX  
 PA (CURA-) CURAGEN CORP.  
 XX  
 PI Leach MD, Shinkets RA;  
 XX  
 WIPI; 2002-106200/14.  
 DR N-PSDB; ABN77683.  
 DR  
 XX Novel human polypeptides and polynucleotides useful for diagnosing,  
 PT preventing and treating cardiovascular disease, neurodegenerative,  
 PT hyperproliferative disorders and disorders related to organ  
 PT transplantation -  
 XX  
 PS Claim 10; Page 1566-1567; 2508pp; English.  
 XX  
 CC Sequences ABP31028-ABP35561 represent 4534 novel human proteins  
 CC designated ORF (open reading frame) 1-4534, and sequences ABN75054-  
 CC ABN79587 represent cDNAs encoding them. The invention also encompasses  
 CC polypeptides at least 80% identical to the ORF1-ORF4534 (collectively  
 CC referred to as ORFX) proteins, polynucleotides at least 85% identical to  
 CC the ORFX nucleic acid sequences, vectors and host cells comprising ORFX  
 CC polynucleotides, the recombinant production of ORFX proteins, antibodies  
 CC specific for ORFX proteins, methods of detecting ORFX polynucleotides and  
 CC polypeptides, methods of screening for modulators of ORFX expression or  
 CC activity, and methods of screening individuals for a predisposition to an  
 CC ORFX-associated disorder. The ORFX proteins of the invention have a wide  
 CC range of biological activities, such as cytokine, cell proliferation,  
 CC cell differentiation, immune modulation, haematopoiesis regulation,  
 CC tissue growth, angiogenesis, activin or inhibin activity, chemotactic/  
 CC chemokinetic activity, haemostatic activity, thrombolytic activity,  
 CC receptor/ligand, antiinflammatory activity, tumour inhibition activity,  
 CC and antiinfective activity, and may also be involved in the determination  
 CC of bodily characteristics, fertility and behaviour. ORFX proteins,  
 CC nucleic acids and antibodies may be used in the treatment of cancers,  
 CC other proliferative disorders such as psoriasis and benign tumours,  
 CC neurological disorders such as epilepsy and Alzheimer's disease,  
 CC cardiovascular diseases, immune system disorders, disorders related to  
 CC organ transplantation, disorders of tissue growth and regeneration,  
 CC diseases such as diabetes mellitus, hypothyroidism, and cholesterol ester  
 CC storage disease, and infectious diseases caused by viral, bacterial,  
 CC fungal and other pathogens. ORFX nucleic acids may also be used as a



CC source of primers and probes, in the detection of ORFX genomic sequences  
 CC or transcripts, in the identification and cloning of homologous  
 CC sequences, in genetic diagnosis, and in forensic biology. The ORFX  
 CC nucleic acids may additionally be used to produce transgenic animals  
 CC which may be useful for studying the function and/or activity of ORFX  
 CC protein, and in drug screening. The ORFX proteins may also be used as  
 CC immunogens to generate specific antibodies, which are useful in the  
 CC diagnosis, treatment and monitoring of ORFX-associated diseases.  
 XX  
 SQ Sequence 108 AA;

Query Match 10.3%; Score 513; DB 23; Length 108;  
 Best Local Similarity 87.0%; Pred. No. 6e-33;  
 Matches 94; Conservative 9; Mismatches 5; Indels 0; Gaps 0;  
 QY 367 VHNRAIRNCPWTVALWSRYLLAMERHGVHDHVISVTFEAKLNAGFTQATDYVEIQAYLD 426  
 DB 1 VHSRAVRNCPWTVALWSRYLLAMERHGLDHTISATFNALSAGFIQATDYVEIQAYLD 60  
 QY 427 YLRRRVDFKQDSSKELELRRAAFTRALEYLKOVEEERFNESEGDPSCVI 474  
 DB 61 YLRRRVDFKQDSSKELELRSMFTRALEYLKOVEEERFSESEGDPSCLI 108

RESULT 14  
 ID ABB59801 standard; Protein; 941 AA.  
 AC ABB59801;  
 XX

DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster polypeptide SEQ ID NO 6195.  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.  
 XX

OS Drosophila melanogaster.  
 XX WO200171042-A2.  
 PN 27-SEP-2001.  
 PD  
 XX 23-MAR-2001; 2001WO-US09231.  
 PF  
 XX 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX (PEKE ) PE CORP NY.  
 PA  
 XX Venter JC, Adams M, Li PWD, Myers EW;  
 PI WPI; 2001-656860/75.  
 DR N-PSDB; ABL03904.  
 XX

PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX

PS Disclosure; SEQ ID.NO 6195; 2lpp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (ABBS7737-ABBT2072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX

SQ Sequence 941 AA;  
 Query Match 8.5%; Score 425; DB 22; Length 941;  
 Best Local Similarity 20.3%; Pred. No. 1.7e-24;  
 Matches 221; Conservative 151; Mismatches 347; Indels 372; Gaps 40;  
 QY 7 TSAPAEABSKAGPKADGEDEYKAARTRRKVLRSRAVAATYKTGMPAWDQOEGYSESD 66  
 DB 87 SSDDPSVEETEGGNAAGR-----GRANDSSSSDDVGVIE--GSELESNSESSDSD 138  
 QY 67 GDEYAMASSAESPGYEYEWEEDEEKEKNOLETERLEEQLSINYDYNCHVDLRLRLLEG 126  
 DB 139 SD-----SDNAGGQNLERSY---QELNALPSKKFAQMV-----LIGIAFKLN 179  
 QY 127 ELTKVRMAROKKSEIFPLTEELWLEWLDHEISMAODGLDREHVDYDLFEKAVDKYICPNIW 186  
 DB 180 DLEKTESSVLELQNLATVPAHVWLKYLKARLVVTTQDDEERKAFEEQCAKALGYYS---- 235  
 QY 187 LEYGQYSGVIGIGQGLKRVSRVFERALS SVGLHMTKGLALMEAYREFESAIVEAARLEK 246  
 DB 236 IPLSEYVNVYLVLDQGNVQN-----HV-----LW-----AKLLA 263  
 QY 247 VHSLFRQLAIPLYDMEATFAEYEE-----WS-----EDPIPESVIQNYNKA 288  
 DB 264 DYDVERPDFGDKLRSLISITITDENEAAAFVEMLOKHCVTWTNCVEORQMIKSYVDKFKQH 323  
 QY 289 LOOLEKYKPYEEALLQAEAPRAEYQAYI--DPE-MKIGDP---ARLQIFERALVENCVLV 343  
 DB 324 LDETRQWDWE-----QKHAHYVDVETLSLDDDLKNAVIRFIFERSVAKFPIV 372  
 QY 344 PDLWIRYSQYLD-----RQLK-----VKDLVLSVHNRAIRNCP 376  
 DB 373 DVLWLSYIEFIOFEGVTVPENEDENEVTAEMVAKRAKRLGKGLRNTLDELANRGVSRHP 432  
 QY 377 WTVALWSRYLLAMERHGVHDHVISVTFEAKLNAGFTQ--ATDYVEIWAQYLDYLRVRV-D 433  
 DB 433 -SVQLNHRFDLMERS--DFELAEVDEEIRL---ILQRTVDMTMDVLFHLDYLAIRN 486  
 QY 434 FKQDSSKELELRRAAFTRALEYLKOVEEERFNESEGDPSVIMONWARIARLNNMQKAR 493  
 DB 487 TNASDEQQVVASLRAAFNHAWE-----ELTVLYGQADTRYEVQLQMAQVEYTLQSGPDNGR 542  
 QY 494 ELWDSIM-TRGNAKYANMLEYNLER-----AHGDTQHCRAKLAHRAVQCTSDYPEHVECV 548  
 DB 543 EIWRQIMGYPGSSIRGLLMLNFAQMESEYNGGHTRDVLRKALSQPV---LENGLMVQEF 599  
 QY 549 LLTMRTEGSLD-----WDIAVQKTETRLARVNEQRMAAEKALVQOEEKEABQRK 602  
 DB 600 FRRYERCYGTYESIAACQALDLPVEYVKPR-SRIKFNQSQSAVPROOKLXPRQOQOQTNRE 658  
 QY 603 RARAEKKALKKKKIRGPEKRGADDEDEKQWGDDEEQPSKRRRVENSIPAAGETQNV 662  
 DB 659 PLNREQ-----RRAHEQOQOQ---QOQKHGIIKSRTE---PSGAT----- 696  
 QY 663 AAGPAGKCAADVPEPPSKQKKAASLKRDMPKVLHDSKSDSITVFFVSNLPYSMOEPDTKL 722  
 DB 697 -----SPSKVK----- 703  
 QY 723 RPLFEACGEVQIRPIFSNRGDFRGYCYVEFKEEKSALEMDRKSVEGRPMFVSPCVD 782  
 DB 704 -----GPANAEAKESN----- 714  
 QY 783 KSKNPDKFVRYSTSLKHKLFISGLPFSCTRKEELEICKAHGTVDKDLRLVTNRACKPKG 842  
 DB 715 -----FKYSPNMEIKFVRNHPACSKRELHELFSFGTINKDLVRLHKLNKQFKG 765  
 QY 843 LAYVEENESQASQAVMKDGMTIKENIIKVAISNPPQRKVPE-KPETRAKAPGGMLLPQ 901  
 DB 766 IAYVEEKPEQAQRAVAGRGCLFKCMNISVAISNPPRPTSAVKPSV-AP----- 815  
 QY 902 TYGARGKGTQLSLPLRALQRPASAAAPQ-----AENGPAAAPAAVAPAAAT----- 946  
 DB 816 -----KRRVPTSLIPTLVROEVAAKKLRLLLPEPGDISTSASVDVAIKREANGEOK 869



Result No.	Score	Query		Length	DB	ID	Description
		Match					
1	272	5.4	414	1	US-07-667-276A-4	Sequence 4, Appli	
2	224.5	4.5	714	2	US-08-990-114-3	Sequence 3, Appli	
3	224.5	4.5	714	4	US-09-241-333-3	Sequence 3, Appli	
4	204	4.1	475	2	US-08-698-407-4	Sequence 4, Appli	
5	204	4.1	475	3	US-09-195-855-4	Sequence 4, Appli	
6	202.5	4.1	1886	4	US-07-938-105-3	Sequence 3, Appli	
7	201	4.0	485	1	US-07-881-075-1	Sequence 1, Appli	
8	201	4.0	485	1	US-08-120-827-1	Sequence 1, Appli	
9	201	4.0	485	1	US-08-478-675-1	Sequence 1, Appli	
10	185.5	3.7	976	4	US-09-104-324B-4	Sequence 4, Appli	
11	185.5	3.7	1939	4	US-09-310-187A-1	Sequence 1, Appli	
12	181	3.6	255	4	US-09-370-838-183	Sequence 183, App	
13	178.5	3.6	323	1	US-07-667-276A-7	Sequence 7, Appli	
14	173.5	3.5	444	1	US-07-881-075-3	Sequence 3, Appli	
15	173.5	3.5	444	1	US-08-120-827-3	Sequence 3, Appli	
16	173.5	3.5	444	1	US-08-478-675-3	Sequence 3, Appli	
17	170	3.4	380	1	US-07-881-075-51	Sequence 51, Appl	
18	170	3.4	380	1	US-08-120-827-51	Sequence 51, Appl	
19	170	3.4	380	1	US-08-478-675-51	Sequence 51, Appl	
20	169.5	3.4	759	1	US-08-676-967-1	Sequence 1, Appli	
21	169.5	3.4	759	1	US-08-676-974-1	Sequence 1, Appli	
22	169.5	3.4	759	2	US-09-098-487-1	Sequence 1, Appli	
23	168.5	3.4	1388	2	US-08-685-576-1	Sequence 1, Appli	
24	168	3.4	652	4	US-09-347-833-2	Sequence 2, Appli	
25	166.5	3.3	2101	1	US-08-468-390-4	Sequence 4, Appli	
26	166.5	3.3	2101	1	US-08-470-950-4	Sequence 4, Appli	
27	166.5	3.3	2101	1	US-08-467-781-4	Sequence 4, Appli	



SEQUENCE CHARACTERISTICS:  
LENGTH: 714 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 128842  
US-09-241-333-3

Query Match 4.5%; Score 224.5; DB 4; Length 714;  
Best Local Similarity 23.5%; Pred. No. 5.8e-09;  
Matches 94; Conservative 67; Mismatches 134; Indels 105; Gaps 16;  
QY 583 AAEKAAALVQOEKAEKARARAE-----KKALKKKKIRGPEKRGAEDE 630  
DB 187 ASEDEDEDEDEDEDEDEDEEAMBITPAKGKAKAPKVPVAKNVAEEDDDDE 246  
QY 631 KEMGDDEEQ-----PSKRRR---VENSIPAAAGETQNV----- 661  
DB 247 EDEDEDEDEDEDEDEDEDEEPPVAPGKRRKKEMTKQKEVPEA-KKQKVEGSEST 305  
QY 662 -----VAAGPACK--CAAVDVPEPPSKQK-----EKA 686  
DB 306 TPFNLFIGNLPNPKVAELKVAISPEFAKNDLAVVDVETGTRKFGYVDFESAEDLEKAL 365  
QY 687 SL-----KRDMPKVLHDSSK--DSITVFVSNLPSYMOEPDKLRPLFEAGGEVVO 734  
DB 366 ELTGLKVFGEIKLEKPKG-RDSKKVRAARTLLAKNLSFNTEDE--LKEVPE---DALE 419  
QY 735 IRPFSNGDPRGYCYVEFEKESALQALEMDR-KSVEGRPMFVSPVDKSKNPDKFKYR 793  
DB 420 IR-LVSDQKSGKIAYIEFKSEADAENLEEKQGAIDGRSVLYTTEKQORBERTGN 478  
QY 794 YSTLEKHKLFISGLPFSCTEKEELEICKAHCYTKDLRLVTRACKPKGLAYVEVENESQ 853  
DB 479 STWGESKTLVLSNLSYSATEETLQEFVEKATFIK---VPONQOKSGKYAFIEFASPED 535  
QY 854 ASQAVKMGDMGTIKENIKVAISNPPQKVPKPTRKAP 893  
DB 536 AKEALNSCKMEIEGTRTIRLELOGP-----RGSPNARSQP 570

RESULT 4  
US-08-698-407-4  
Sequence 4, Application US/08698407  
Patent No. 5856128  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Au-Young, Janice  
APPLICANT: Hawkins, Phillip R.  
APPLICANT: Hillman, Jennifer L.  
TITLE OF INVENTION: A NOVEL HUMAN NUCLEIC ACID BINDING PROTEIN  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: U.S.  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/698,407  
FILING DATE: Filed Herewith  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0116 US

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 475 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 267188  
US-08-698-407-4

Query Match 4.1%; Score 204; DB 2; Length 475;  
Best Local Similarity 22.9%; Pred. No. 1.2e-07;  
Matches 85; Conservative 68; Mismatches 174; Indels 44; Gaps 12;  
QY 559 LEDMD-IAVQKTETRLARVNEQRMKAKEAALVQOEKAEKQAR-KRARAEEKALKKKKK 616  
DB 1 MSDDFEFERQNLNENKQERDKNHRKRSRSDRKRKRSRRDRNRDQSRASDRRR 60  
QY 617 IRGPKRGAEDEDEKEMGDDEEQPSKRRRVENSIPAAE--TQNVVAAGPAGCAAVD 674  
DB 61 RSKPLTRGAKKEHGLIRSPRHEKKKKVKRYWDPVPPGFHEITPMQYKAMQAAGQIPATA 120  
QY 675 VEPPSKQKAEASLAKRMPKVLHDSKDSITVFVSNLPSYMOE-----PDKLRPLF 726  
DB 121 L-LPTMTDGLAVTTPVPVVGSGQTRQARLLYGVNIPFGITEEAMDMDFNAQMLGGLT 179  
QY 727 EACGE---VVQIRPISNRGDFRGYCYVEFEKESALQALEMDRKSVEG-----RPMFV 777  
DB 180 QAPGNPLAVQI-----NQDNFALEFRSVDETTQAMAFDGIIFQOSLKIRPHDY 232  
QY 778 SPVDKSKNPDKFKYRSTSL---EKHKLFISGLPFSCTEKEELEICKAHGVKDLRLVT 834  
DB 233 QPLPKMSENPSYVPGVSTVVPVPSAHKLFIFGLPLNYLNDQVKELLTSFGPLKAFNLVK 292  
QY 835 NRA-CKPKGLAYVEVENESQASQAVKMGDMGTIKENIKV-----ALSNPPQKRV 883  
DB 293 DSATGLSGYAFCEYVDINVTQATAGLNGMQLGDKLLVQRASVGAKNATLVSPPS-TI 351  
QY 884 PEKPTRKAPG 894  
DB 352 NQTPVTLQVPG 362

RESULT 5  
US-09-195-855-4  
Sequence 4, Application US/09195855  
Patent No. 6015788  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Au-Young, Janice  
APPLICANT: Hawkins, Phillip R.  
APPLICANT: Hillman, Jennifer L.  
TITLE OF INVENTION: A NOVEL HUMAN NUCLEIC ACID BINDING PROTEIN  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: U.S.  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/195,855

FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/698,407  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0116 US  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 475 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 267188  
US-09-195-855-4

Query Match 4.1%; Score 204; DB 3; Length 475;  
Best Local Similarity 22.9%; Pred. No. 1.2e-07;  
Matches 85; Conservative 68; Mismatches 174; Indels 44; Gaps 12;  
QY 559 LEDWD-IYVQETRLARVNEORMKAAEKAALVQEEBKAQR-KRAAEKALKKKK 616  
DB 1 MSDDFEFQNLNENQERDKENHRKRSRDRKRKRSDRRNRDQSRSDRRR 60  
QY 617 IRGPKRGADDEKWDGDEEOPSKRRRVENSIPAAGE--TONVEAAGPAGCAAVD 674  
DB 61 RSKPLTRGAKKEHGLIRSPREKKKKVKYWDVPPGFEHITPMQYKAMQAAGQIPATA 120  
QY 675 VEPSPQKEKAASLRDKPKVLHDSKDSITVFSNLPYSMQE-----PDTKLRLPLF 726  
DB 121 L-LPTMTDGLAVTPTFPVVGVSQMTQARRLYVGNIPFGITEAMMDFNMQRLGGIT 179  
QY 727 EACGE---VVOIRPFSNRGDFRGYCYVEFEKESALQALEMDRKSVEG-----RPMFV 777  
DB 180 QAPGNVLAVQI-----NODKNFALFEFRSVDETTQAMAFDGIIFQOQSLKIRPHDY 232  
QY 778 SPVCKSNKNDPKVRYSTSL---EKHKLFISGLPFSCTKEELEICKAHGTVDKDLRLVT 834  
DB 233 QPLPGMSENPVPGVSTVVPDPSAHKLFIGGLPNYLNDDQVKELLTSGPLKAFNLVK 292  
QY 835 NRA-GKPKGLAYVEENESQASQAVMKMDGMTIKENIKV-----AISNPPQKV 883  
DB 293 DSATGLSKGYAFCEYVDINVTQATAGLNGMLQDKKLLVQRASVGAKNATLVSPPS-TI 351  
QY 884 PEKPTRKAPG 894  
DB 352 NQTPVTLQVP 362

RESULT 6.  
US-09-938-105-3  
Sequence 3, Application US/08938105  
Patent No. 6353151  
GENERAL INFORMATION:  
APPLICANT: Leinwand, Leslie A.  
APPLICANT: Vikstrom, Karen L.  
TITLE OF INVENTION: TRANSGENIC MODEL FOR HEART FAILURE  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross P.C.  
STREET: 1700 Lincoln St., Suite 3500  
CITY: Denver  
STATE: CO  
COUNTRY: U.S.A.  
ZIP: 80203  
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/938,105  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Crook, Wannel M.  
REGISTRATION NUMBER: 31,071  
REFERENCE/DOCKET NUMBER: 3595-4  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1886 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-938-105-3  
Query Match 4.1%; Score 202.5; DB 4; Length 1886;  
Best Local Similarity 19.6%; Pred. No. 1.2e-06;  
Matches 195; Conservative 154; Mismatches 342; Indels 303; Gaps 46;  
QY 16 SKAGPKADGEDEVAARTRRRKVLGRAAAATYKTMGPADWQOEGVSESDGDEVAMASS 75  
DB 970 TSKVKLEQVDDLSGSEQKKVRMDLERARKLEGDL-KUTOSIMDLENDKILQLEBK 1028  
QY 76 AESSPEYEWYDEEENKQLEIE-----RLSEQLSINVYDYNCHVDLIRLRLEGETLK 130  
DB 1029 LKKK-----EPDISQNSKIEDEQALALQLOKLEK-----QARIEELEEA 1073  
QY 131 VMAKQKSEI-FPITLWELWLEHDEISMAODGLDREHVYDLFEKAVKDYICPNWLEY 189  
DB 1074 ERTARAKVEKLSDTREL-----EISERLEA----- 1102  
QY 190 GQVSGGIGQKGLKRVSRFALSSVGLHMTKGLALWEAYREFESAIVEAARLEKVHS 249  
DB 1103 -----GG-----ATSVQIENK-----KRAEFQK----- 1122  
QY 250 LFRROLAIPLYDMEATFAEYEWSEDPPE--SVIONYNKALQOLEKYK----- 296  
DB 1123 -MRDLEATLQHEATAAARLKHADSVAEELGEQIDNLRVQKLEKESEKLELDDVT 1181  
QY 297 PYEEALLQEA-----PRLAEYOAYIDPEMKI-----GDPARTQ 330  
DB 1182 SHMEQIIKAKANLEKVSRTLEDQAN-EYRVKLEEAQRSINDFTTORAKLQTEGELAR-Q 1239  
QY 331 LIFERALVENCULVPDLW-----IRYSOYLD---RQL-----KVKDLVLSVHNRAIRNCBW 377  
DB 1240 LEEKEALI-----WQTRGKLSYTOQMEDLQKLEEGKAKNALAHALQSARHDCD- 1290  
QY 378 TVALMSRYLLAMERHGVHDQVISVTFEAKNALNAGFIQATDYVEIWOAYLDYLRVRVDFKOD 437  
DB 1291 --LLREQYEEEMAEALQRVLS-----KANSEVAQW-----RTKYETD 1327  
QY 438 SSKLEELRAAFTRALEYLKQVEERFNEGDPSCVIMONWARIARLNNMQ----- 490  
DB 1328 AIQRTTEELEEAQKLAQRL-QDAEEAV-EAVNAKCSLE---KTHRLQNETEDLMVDVVE 1382  
QY 491 -----KARELWDSIMTRGNAKYANMMWLEYNLERAHGDTQCHRKALHRAVQCTSD 540  
DB 1383 RSNAALADKKQNFDKILAEWKQKYESQSE---LESSQKEARSLTELEKLNAYEE 1439  
QY 541 YPEHVCEVLLTWERTEGSLED--WDIAVQ-----KTETRLARVNEQ-RMKAAEKAALVQ 592  
DB 1440 SLEH-----LETFKRKNLQEEISDLTQEGGGKGVHLEKIRKQLEVEKLEQSALEE 1495  
QY 593 QEEE-KAEQRKARAEKALKKKKIRGPEKGADEDEKEWG-----DD 636

Db 1496 AEASLEHEGKILRALEFNOIKAEI-----ERKLAKEDEMBQAKRNHLRVVDSLQTSIDA 1552  
 QY 637 EEEQPSKRRRYVENSIPAAGETONVEAAGPAGKCAAVDVEPPSKQKKAASLKRDMPKV- 695  
 Db 1553 ETRSNEALRYKKM--EGDLNEMEIQLSQNRITAS---EAKHLKNAQAHLKDTQLQD 1607  
 QY 696 ----LHOSKSDITVFSNLPYSMOEPTKRLPLFEAGGEVYQIRPFSNRGDFRGYCV 751  
 Db 1608 DAVRANDLKENIAI-----VERRNTLLQ-----AELELRVVEQTERSRLAEQ 1653  
 QY 752 EFKEKSALQALEMDRKSVEGRPMFVSPVDKSNPKVPFRYSTSL-----EK 800  
 Db 1654 ELIETSERVQLLHQSNTSL-----INOKKMDADLSQLQTEVEEAVOEACRNBK 1703  
 QY 801 HKLFTSLGPFSCFEELEECIAH-----GTVKDLRLVTNRA-----GRPKGLAY 845  
 Db 1704 AKKAITDAMMAEELKREODTSAHLERMKKNMEQTIKDLQHLDEAEQIALKGGKKLOK 1763  
 QY 846 V-----EYENESQASQA--VMKMDGTIKENI 872  
 Db 1764 LEARVRELENELEAEQKRNAESVKGMRKSERRIK 1797

RESULT 7

US-07-881-075-1  
 ; Sequence 1, Application US/07881075  
 ; Patent No. 544149  
 ; GENERAL INFORMATION:  
 ; APPLICANT: KEENE, JACK D.  
 ; APPLICANT: KING, PETER H.  
 ; APPLICANT: LEVINE, TODD  
 ; TITLE OF INVENTION: METHODS AND COMPOSITIONS USEFUL IN THE  
 ; TITLE OF INVENTION: RECOGNITION, BINDING AND EXPRESSION OF RIBONUCLEIC ACIDS  
 ; TITLE OF INVENTION: INVOLVED IN CELL GROWTH, NEOPLASIA AND IMMUNOREGULATION  
 ; NUMBER OF SEQUENCES: 51  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
 ; ADDRESSEE: P.C.  
 ; STREET: 1755 Jefferson Davis Highway, Fourth Floor  
 ; CITY: Arlington  
 ; STATE: Virginia  
 ; COUNTRY: U.S.A.  
 ; ZIP: 22202  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/07/881,075  
 ; FILING DATE: 19920511  
 ; CLASSIFICATION: 530  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Oblon, No. 544149man F.  
 ; REGISTRATION NUMBER: 24,618  
 ; REFERENCE/DOCKET NUMBER: 714-154-0  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (703)521-4500  
 ; TELEFAX: (703)486-2347  
 ; TELEX: 248855 OPAT UR  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 485 amino acids  
 ; TYPE: AMINO ACID  
 ; TOPOLOGY: unknown  
 ; MOLECULE TYPE: protein  
 ; US-07-881-075-1

Query Match 4.0%; Score 201; DB 1; Length 485;  
 Best Local Similarity 20.7%; Pred. No. 2.1e-07;  
 Matches 91; Conservative 73; Mismatches 136; Indels 140; Gaps 16;  
 QY 564 IAVQKTETRLA-----RVNEQRMKAAEKAALVQOEEKEAEQKRAEKKALKKK 614

Db 64 LQVQOQOQQAVAAAAAVTQLOQQOQAVVAQVQOQOQAAAAVQAAVQAAVVPQ 123  
 QY 615 KKIRCEPKRGADEDEKEDGDEEBQPSKRRRVENSIPAAGETONVEAAGPAGKCAAVD 674  
 Db 124 PQOQOPNTNG-----NAGSGSQN-----GSG----- 145  
 QY 675 VEPFSKQKKAASLRKDPKVLHDSKSDITVFSNLPYSMOEPTKRLPLFEAGGEVYQ 734  
 Db 146 -----STETRTNLIVNLPQTWTEDE--IRSLFSSVGEIES 179  
 QY 735 IR-----PIFSNRGDRPGYCYVEFEKKEKSALQA-----LEMDRKSVEGRP 774  
 Db 180 VKLIRKDSQVYIDPLNPOAPSKGSLGXFVYVRPQDAEQAVNVLNGLRLQNKTI--KV 237  
 QY 775 MFVSPVDKSNPKVPFRYSTSLKHLFTSLGSLFSCFKELEECIAHGTVKDLRLVT 834  
 Db 238 SFARSSDAIKGAN-----LYVSGLPKTMQTELEAIFAPFGAITTSRIQ 283  
 QY 835 NRAG--KPKGLAYVEYENESQASQAVMKMDGTIKE--NIKVAISNPP-----QRKV 883  
 Db 284 NAGNDTQTKGVGFIRFDKREERATRAIALNGTTPSSCTDPIVVKFSNTPGSTSKIQPOL 343  
 QY 884 PE--KPETRKAPGGPMLLPQTVGARGKGTQLS-----LLPLRALQRPSSAAAPQAN 932  
 Db 344 PAFLNPQLVRRIGGAMHTPV-----NKGLARFSPMAGDMLDVMPLNGLGAAAAAATTLAS 398  
 QY 933 GPAAAPAV-----AAPAATEA 948  
 Db 399 GPGGAYPIFYINLAPETEEA 418

RESULT 8

US-08-120-827-1  
 ; Sequence 1, Application US/08120827  
 ; Patent No. 5525495  
 ; GENERAL INFORMATION:  
 ; APPLICANT: KEENE, JACK D.  
 ; APPLICANT: KING, PETER H.  
 ; APPLICANT: LEVINE, TODD  
 ; TITLE OF INVENTION: METHODS AND COMPOSITIONS USEFUL IN THE  
 ; TITLE OF INVENTION: RECOGNITION, BINDING AND EXPRESSION OF RIBONUCLEIC ACIDS  
 ; TITLE OF INVENTION: INVOLVED IN CELL GROWTH, NEOPLASIA AND IMMUNOREGULATION  
 ; NUMBER OF SEQUENCES: 101  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
 ; ADDRESSEE: P.C.  
 ; STREET: 1755 Jefferson Davis Highway, Fourth Floor  
 ; CITY: Arlington  
 ; STATE: Virginia  
 ; COUNTRY: U.S.A.  
 ; ZIP: 22202  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/120,827  
 ; FILING DATE: 15-SEP-1993  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Oblon, No. 5525495man F.  
 ; REGISTRATION NUMBER: 24,618  
 ; REFERENCE/DOCKET NUMBER: 714-158-0 CIP  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (703)413-3000  
 ; TELEFAX: (703)413-2220  
 ; TELEX: 248855 OPAT UR  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 485 amino acids  
 ; TYPE: amino acid







QY 396 ----HOVSVTFEKA-INAGFIQATDYVEIWOAYLDYLRVRVDFKODSSKELEELRAAFTR 451  
Db 1344 LLREQVEETEAKAELQRLVLSKANSEVAQW-----RTKYETDAIQRTTEELBEAKKK 1394  
QY 452 ALEYLKQVEERPNESGDSVIMQWARIARLNNMQ-----KARE 494  
Db 1395 LAQRL-ODAEAEV-EAVNAKCSLE---KTKHRLQNEIEDLMVDVERSNAALDKKOR 1449  
QY 495 LWDSIMTRGNKAYNMWLEYINLERAHGDTQHCRCALHRAVQCTSDYPHVCEVLLTMR 554  
Db 1450 NFDKILAEWKQYEEQSE---LESSQKARSLSLTFKLKNAYEBSLEH----LETFRK 1502  
QY 555 TEGSLD--WDIAVQ-----KTETRLARVNEQ-RMKAKEAALVQOEEE-KAQQRKRAR 605  
Db 1503 ENKNLQEIISDLTEQLGEGGKNVHELEKVRKQLEVEKLELQSALEAEASLEHEGKIILR 1562  
QY 606 AEKKALKKKKIRGPKRGADEDEKEWG-----DDEEQPSKRRRVENS 650  
Db 1563 AOLFEHQIAEI---ERKLAERKDEEMEQAKRHRVVDLSQTSLSDAETRSREVLVRKKK 1619  
QY 651 IPAAGETONVEAAGPAGCAADVPEPPSKQKEKAASLRDMPKVL-----HDSSKDSI 704  
Db 1620 M--EGDLNEMETOLSHANMAA---EAKQVKVLSQSLKDTQIQIQLDDAVRANDDLKENI 1673  
QY 705 TVFVSNLPSYMOEPTKRLPLFEACGEVVOIRPIFSNRGDFRGYCYVEFEKESALQALE 764  
Db 1674 AIV-----ERRNLLQA--ELEELRAVVEQTERSRLAEQELIETSERVOLLH 1719  
QY 765 MDRKS-VEGRPMFVSPVDKSNPDFKVPYRSTSLKHKLFISGLPFCSTKEELEICEKA 823  
Db 1720 SONTSLINOKKMSDLTOLQSEVEAEVQECRNEAEKAKAITDAMMAEELKKEQDTS 1779  
QY 824 H-----GTVKDLR-----LVTNRAGPK-----GLAYVEYENESQASOAVMKM 861  
Db 1780 HLEMRKKNMEQTIKDLQHLRDAEQIALKGGKKQKQKLEARVRELEGELEAEQKRNESV 1839  
QY 862 DQMTKENTIK 872  
Db 1840 KGMRSERRIK 1850

## RESULT 12

US-09-370-838-183  
; Sequence 183, Application US/09370838  
; Patent No. 6444425

## GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Mohamath, Roadon  
; APPLICANT: Secrist, Heather  
; TITLE OF INVENTION: COMPOUNDS FOR THERAPY AND DIAGNOSIS OF  
; LUNG CANCER AND METHODS FOR THEIR USE  
; FILE REFERENCE: 210121.475C1  
; CURRENT APPLICATION NUMBER: US/09/370,838  
; CURRENT FILING DATE: 1999-08-09  
; EARLIER APPLICATION NUMBER: US 09/285,323  
; EARLIER FILING DATE: 1999-04-02  
; NUMBER OF SEQ ID NOS: 289  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 183  
; LENGTH: 255  
; TYPE: PRT  
; ORGANISM: Homo sapien

## US-09-370-838-183

Query Match 3.6%; Score 181; DB 4; Length 255;

Best Local Similarity 23.3%; Pred. No. 2.9e-06;

Matches 66; Conservative 51; Mismatches 104; Indels 62; Gaps 10;

QY 653 AAGETONVEVAA-----GPGAKCAADVPEPPSKQKEKAASLRD 691

Db 2 AAGVAAAEVAAETEPKMEESGAPVPSGNGAPGPKG-----EERTQNEKRKEKNIRG 56

QY 692 MPKV-LHDSKDSITVFNLSNPSMOEPTDKLRPL-----FEACGEVVOIRPIFSNRGDFR 746  
Db 57 GNRFEPSNPTKRYRAFITNIPF-----DVKQSLKDLVKEKVGVEVYVELLMDABGKSR 111  
QY 747 GYCYVEFEKESALQALE-MDRKSVEGRPMFVSPVDKSNPDFKVPYRSTSLK-----800  
Db 112 GCADVFEKMEESKMAAEVNLKSHLSGRPLKV-----KEDPDGEHAR--RAMQKAGRLG 163  
QY 801 HKLFISGLPFCSTKEELEICEKAHGTVDKLRVLTNRAGPKGLAYVEYENESQASOAVMK 860  
Db 164 STVFVANLDYKVGWKKLKEVFSMAGVVVRADILEDKGKSRGIGIVTFEQSIEAQAQISM 223  
QY 861 MDGMTIKENIIKVAISNPPQRRVPEKPETRKAPGGPMLLPQY 903  
Db 224 FNGQLLFD-----RPMHVKMDERALPKGDFPPPERH 254

## RESULT 13

US-07-667-276A-7

; Sequence 7, Application US/07667276A

; Patent No. 5470971

## GENERAL INFORMATION:

; APPLICANT: Kondo, Keiji

; APPLICANT: Inouye, Masayori

; TITLE OF INVENTION: STRESS-INDUCED PROTEINS, GENES CODING

; THEREFOR, TRANSFORMED CELLS OF ORGANISMS, METHODS AND

; TITLE OF INVENTION: APPLICATIONS

; NUMBER OF SEQUENCES: 10

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Weiser & Associates

; STREET: 230 S. Fifteenth Street, Suite 500

; CITY: Philadelphia

; STATE: PA

; COUNTRY: USA

; ZIP: 19102

## COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/667,276A

; FILING DATE: 11-MAR-1991

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Weiser, Gerard J.

; REGISTRATION NUMBER: 19,763

; REFERENCE/DOCKET NUMBER: 377,5351P

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 215-875-8383

; TELEFAX: 215-875-8394

; INFORMATION FOR SEQ ID NO: 7:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 323 amino acids

; TYPE: amino acid

; TOPOLOGY: unknown

; MOLECULE TYPE: protein

US-07-667-276A-7

## Query Match

3.6%; Score 178.5; DB 1; Length 323;

Best Local Similarity 26.1%; Pred. No. 6.4e-06;

Matches 61; Conservative 42; Mismatches 88; Indels 43; Gaps 9;

QY 660 VEVAAGPAGKCAADVPEPPSKQKEKAASL-----KRDMPKVLHDSKDSITVFPVS 709

Db 34 VDVIRIGMTRKFGYVDFE-SAEDEKALELTGLKVFNGEIKLEKPK--GKDSKARTLLAK 90

QY 710 NLPSYMOEPTKRLPLFEACGEVVOIRPIFSNRGDFRGYCYVEFEKESALQALEMDR-K 768

Db 91 NLPYKVTODE--LKEVFEDAAEI-----RLVSKDGSKGIATIEFTKTEADAEKTFEKOQT 144

QY 769 SVEGRPMFVSPVDKSNPDFKVPYRSTSLKHK-----LFTISGLPFCSTKEELEICEKAH 824

Db 145 EIDGR-----SISLYTGERKQNSKTLVLSNYSATELTQEVFEKA 188  
QY 825 GTVKDLRLVTNRAGPKGLAYVEVENESQASQAVKMDGMTIKENIKVAISNP 878  
Db 189 TFIK---VPQNGKSKGYAIEFASPEDAKALNSCNKREIEGRAIRLEQGP 239

RESULT 14  
US-07-881-075-3  
; Sequence 3, Application US/07881075  
; Patent No. 5441419  
; GENERAL INFORMATION:  
; APPLICANT: KEENE, JACK D.  
; APPLICANT: KING, PETER H.  
; APPLICANT: LEVINE, TODD  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS USEFUL IN THE  
; TITLE OF INVENTION: RECOGNITION, BINDING AND EXPRESSION OF RIBONUCLEIC ACIDS  
; TITLE OF INVENTION: INVOLVED IN CELL GROWTH, NEOPLASIA AND IMMUNOREGULATION  
; NUMBER OF SEQUENCES: 51  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,  
; STREET: 1755 Jefferson Davis Highway, Fourth Floor  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/881,075  
; FILING DATE: 19920511  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Oblon, No. 5441419man F.  
; REGISTRATION NUMBER: 24,618  
; REFERENCE/DOCKET NUMBER: 714-154-0  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)521-4500  
; TELEFAX: (703)486-2347  
; TELEX: 248855 OPAT UR  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 444 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
US-07-881-075-3

Query Match 3.5%; Score 173.5; DB 1; Length 444;  
Best Local Similarity 23.1%; Pred. No. 2.5e-05;  
Matches 81; Conservative 55; Mismatches 122; Indels 93; Gaps 14;

QY 641 PSKRRVENS-IPAAGETQNVAAAGPAGKC--AAVDVPPSKQKEKAASLRDMPKVLH 697  
Db 34 PATANNVTSQAQTNGGTATTAAAGAGSTTNAAVG-----QATANNAASNNNNNNNTN 89  
QY 698 DSSKDSIT-----VFVSNLPYSMQEPDTKLRPLFEACGEVYQIRPFSN-R 742  
Db 90 NNNNNNATANNNNNEPDKTNLIVNLPQTSQDE--IRSLFVSFGEVESCRLIRDKVT 147  
QY 743 GDFRCYCYVEFKEEK-----SALOALEMDRKSVE---GRPMFVSPCVDKSKNPDKVFR 793  
Db 148 GOSLGYGVXYVKQDAEKAINALNGLRLQNTIKVSIARPS-----189  
QY 794 YSTLEKHKLFISGLPFSCFCTKEELEICKAHGTVKDLRLVTNR-----AGPKGLAYVE 847  
Db 190 -SESIKANLYVSGLPKNTQSDLESFLSPYGIKITSRLICDNTITDEHAAGLSKGVGFIR 248

QY 848 YENESQASQAVKMDGMTIKENIKVAISNPQKVPKPEKTRKAPGPGMLLPQTYGARG 907  
Db 249 FQDRPEADRAIKELNGTTPKNS-----TEPTTVKFANNPSSNKN-----287  
QY 908 KGRTOLSLLRALQRPASAAQOENG-----PAAADPAVAAPATEAPKMSNA 954  
Db 288 -----SMOPLA-----AYIAPONTGRRAPFANAAAGAAAAAAAIHPNA 328

RESULT 15  
US-08-120-827-3  
; Sequence 3, Application US/08120827  
; Patent No. 5525495  
; GENERAL INFORMATION:  
; APPLICANT: KEENE, JACK D.  
; APPLICANT: KING, PETER H.  
; APPLICANT: LEVINE, TODD  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS USEFUL IN THE  
; TITLE OF INVENTION: RECOGNITION, BINDING AND EXPRESSION OF RIBONUCLEIC ACIDS  
; TITLE OF INVENTION: INVOLVED IN CELL GROWTH, NEOPLASIA AND IMMUNOREGULATION  
; NUMBER OF SEQUENCES: 101  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,  
; STREET: 1755 Jefferson Davis Highway, Fourth Floor  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/120,827  
; FILING DATE: 15-SEP-1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Oblon, No. 5525495man F.  
; REGISTRATION NUMBER: 24,618  
; REFERENCE/DOCKET NUMBER: 714-158-0 CIP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)413-3000  
; TELEFAX: (703)413-2220  
; TELEX: 248855 OPAT UR  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 444 amino acids  
; TYPE: amino acid  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
US-08-120-827-3

Query Match 3.5%; Score 173.5; DB 1; Length 444;  
Best Local Similarity 23.1%; Pred. No. 2.5e-05;  
Matches 81; Conservative 55; Mismatches 122; Indels 93; Gaps 14;

QY 641 PSKRRVENS-IPAAGETQNVAAAGPAGKC--AAVDVPPSKQKEKAASLRDMPKVLH 697  
Db 34 PATANNVTSQAQTNGGTATTAAAGAGSTTNAAVG-----QATANNAASNNNNNNNTN 89  
QY 698 DSSKDSIT-----VFVSNLPYSMQEPDTKLRPLFEACGEVYQIRPFSN-R 742  
Db 90 NNNNNNATANNNNNEPDKTNLIVNLPQTSQDE--IRSLFVSFGEVESCRLIRDKVT 147  
QY 743 GDFRCYCYVEFKEEK-----SALOALEMDRKSVE---GRPMFVSPCVDKSKNPDKVFR 793  
Db 148 GOSLGYGVXYVKQDAEKAINALNGLRLQNTIKVSIARPS-----189  
QY 794 YSTLEKHKLFISGLPFSCFCTKEELEICKAHGTVKDLRLVTNR-----AGPKGLAYVE 847  
Db 190 -SESIKANLYVSGLPKNTQSDLESFLSPYGIKITSRLICDNTITDEHAAGLSKGVGFIR 248

Search completed: June 18, 2003, 13:52:47  
Job time : 30 secs

~~THIS PAGE BLANK~~ (USPTO)

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 18, 2003, 13:51:48 ; Search time 55 Seconds  
(without alignments)  
1894.603 Million cell updates/sec

Title: US-09-763-985A-2  
Perfect score: 4994  
Sequence: 1 MATAETASPEAESKAGP.....AATEAPKMSNADFAKFLRK 963.

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 41779 seqs, 108206813 residues  
Total number of hits satisfying chosen parameters: 41779

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_AA:\*  
1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	4989	99.9	1179	10	US-09-821-883-29
2	2918	58.4	578	10	US-09-925-300-1496
3	526	10.5	104	10	US-09-864-761-36026
4	231	4.6	687	9	US-09-969-384-27
5	225	4.5	708	10	US-09-925-301-1012
6	224.5	4.5	714	10	US-09-978-242-3
7	217	4.3	41	10	US-09-864-761-34622
8	208.5	4.2	1938	9	US-10-171-311-164
9	208.5	4.2	1945	9	US-09-927-597-2
10	208.5	4.2	1972	9	US-10-171-311-162
11	208.5	4.2	1979	9	US-09-927-597-4
12	206.5	4.1	494	10	US-09-833-790-234
13	199.5	4.0	852	10	US-09-752-639-153
14	199.5	4.0	852	10	US-09-984-198-153
15	199	4.0	687	9	US-09-969-384-16
16	194	3.9	446	10	US-09-925-300-1554
17	188.5	3.8	1203	9	US-10-097-340-43
18	186.5	3.7	482	9	US-10-153-668-320
19	184.5	3.7	486	9	US-10-153-668-432

ALIGNMENTS

RESULT 1		US-09-821-883-29		Sequence 29, Application US/09821883		Sequence 41, Appl	
		; Patent No. US20020061310A1		; GENERAL INFORMATION:		Sequence 260, App	
		; APPLICANT: Laus, Reiner		; APPLICANT: Gradvic, Damir		Sequence 183, App	
		; APPLICANT: Vidovic, Thomas		; TITLE OF INVENTION: Cell-Based Immunotherapy		Sequence 183, App	
		; TITLE OF INVENTION: Compositions and Methods for Dendritic		; FILE REFERENCE: 7636-0022.30		Sequence 29, Appl	
		; CURRENT FILING DATE: 2001-03-30		; PRIOR APPLICATION NUMBER: US 60/193,504		Sequence 5, Appl1	
		; PRIOR FILING DATE: 2000-03-30		; NUMBER OF SEQ ID NOS: 30		Sequence 10, Appl	
		; SOFTWARE: FastSeq for Windows Version 4.0		; SEQ ID NO 29		Sequence 27, Appl	
		; LENGTH: 1179		; TYPE: PRT		Sequence 388, App	
		; ORGANISM: Artificial Sequence		; FEATURE:		Sequence 392, App	
		; OTHER INFORMATION: SART-3-IC		US-09-821-883-29		Sequence 272, App	
		Query Match		99.9%; Score 4989; DB 10; Length 1179;		Sequence 4611, Ap	
		Best Local Similarity		100.0%; Pred. No. 2.7e-309; Mismatches 0; Indels 0; Gaps 0;		Sequence 16, Appl	
		Matches 962; Conservative 0;				Sequence 16, Appl	
QY		1		MATAETASPEAESKAGPKADGEDEVKAAARTRRKVLRAVAATAATYKTGMPADQDEE 60		Sequence 16, Appl	
Db		1		MATAETASPEAESKAGPKADGEDEVKAAARTRRKVLRAVAATAATYKTGMPADQDEE 60		Sequence 16, Appl	
QY		61		GVSESDGDEYAMASSAESPGYEYDEEEKNOLETERLEEQLSINVDYNCHVDLIR 120		Sequence 16, Appl	
Db		61		GVSESDGDEYAMASSAESPGYEYDEEEKNOLETERLEEQLSINVDYNCHVDLIR 120		Sequence 16, Appl	
QY		121		LRLRGEELTKVMAQKMSIEIPPLFEELWHLWHDSEISMAQDGLDREHYVDLFEKAVKDY 180		Sequence 16, Appl	
Db		121		LRLRGEELTKVMAQKMSIEIPPLFEELWHLWHDSEISMAQDGLDREHYVDLFEKAVKDY 180		Sequence 16, Appl	
QY		181		ICPNIWLEYGQYSVGGIGQKGLGKVRSVFERALSSVGLHMTKGLALWEAYREFESAIVE 240		Sequence 16, Appl	

Db 181 ICNINILEYGOYVGGIGQGGLEKRVSVFERALSSVGLHMTKGLALWEAYREFESAIVE 240  
QY 241 AARLEKVHSLFRQLAIPLYDMEATFAEYEEWSEDPJPESVIONYNKALQOLEKYPYEE 300  
Db 241 AARLEKVHSLFRQLAIPLYDMEATFAEYEEWSEDPJPESVIONYNKALQOLEKYPYEE 300  
QY 301 ALLOAEAPRLAEOYAYIDFEMKIGDPARQIOLIFERALVENCNCLVDPDLWIRYSQYLDRLQKV 360  
Db 301 ALLOAEAPRLAEOYAYIDFEMKIGDPARQIOLIFERALVENCNCLVDPDLWIRYSQYLDRLQKV 360  
QY 361 KDILVSVHNRAINCPTWVALMSRYLLAMERHGVHDQVTSVTEKALNAGFIQATDYVEI 420  
Db 361 KDILVSVHNRAINCPTWVALMSRYLLAMERHGVHDQVTSVTEKALNAGFIQATDYVEI 420  
QY 421 WQAYLDYLRVRVDFKODSSKELEELRAAFTRALEYLKQVEEERFNSGDPSCVIMONWAR 480  
Db 421 WQAYLDYLRVRVDFKODSSKELEELRAAFTRALEYLKQVEEERFNSGDPSCVIMONWAR 480  
QY 481 IEARLCNNMOKARELWDSIMTRGNAKYANNMWLEYYNLERAHGDTQHCRCALHRAVQCTSD 540  
Db 481 IEARLCNNMOKARELWDSIMTRGNAKYANNMWLEYYNLERAHGDTQHCRCALHRAVQCTSD 540  
QY 541 YPEHVCEVLLTMERTEGSLEDWDIAVQKTETRLARVNEORMKAAEKEAALVQOEESKAEQ 600  
Db 541 YPEHVCEVLLTMERTEGSLEDWDIAVQKTETRLARVNEORMKAAEKEAALVQOEESKAEQ 600  
QY 601 RKRAAEKALKKKKIRGPEKRGADDEDEKQWGDDEEOPSKRRRVENSIPAAGETQNV 660  
Db 601 RKRAAEKALKKKKIRGPEKRGADDEDEKQWGDDEEOPSKRRRVENSIPAAGETQNV 660  
QY 661 EVAAGPAGCAAVDVPEPPSKQKEKAASLRDMPKVLHDSKSDSITVFSNLPYSMQEPDT 720  
Db 661 EVAAGPAGCAAVDVPEPPSKQKEKAASLRDMPKVLHDSKSDSITVFSNLPYSMQEPDT 720  
QY 721 KLRLPEACEGVVQIRPIFSNRGDFRGYCYVEFEKEKSALQALEMDRKSVEGRPMFVSPC 780  
Db 721 KLRLPEACEGVVQIRPIFSNRGDFRGYCYVEFEKEKSALQALEMDRKSVEGRPMFVSPC 780  
QY 781 VDKSKNPDKFVFRYSTSLEKHKLFISGLPFSCCTKELEICAHGTVKDLRLVTNRAGKP 840  
Db 781 VDKSKNPDKFVFRYSTSLEKHKLFISGLPFSCCTKELEICAHGTVKDLRLVTNRAGKP 840  
QY 841 KGLAYVEYENESQASQAVMKMDGWTIKENIIKVAISNPPQKVPKPTRKAPGGPMLLP 900  
Db 841 KGLAYVEYENESQASQAVMKMDGWTIKENIIKVAISNPPQKVPKPTRKAPGGPMLLP 900  
QY 901 QTYGARGKGTQSLPRALQPSAAPOAENGPAAPAAVAPAAPEAPKMSNADFAKLF 960  
Db 901 QTYGARGKGTQSLPRALQPSAAPOAENGPAAPAAVAPAAPEAPKMSNADFAKLF 960  
QY 961 LR 962  
Db 961 LR 962

RESULT 2

US-09-925-300-1496  
; Sequence 1496, Application US/09925300  
; Patent No. US20020151681A1  
; GENERAL INFORMATION:  
; APPLICANT: Craig Rosen,  
; APPLICANT: Steve Ruben  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA101  
; CURRENT APPLICATION NUMBER: US/09/925,300  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05988  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 1890  
; SOFTWARE: Patentin Ver. 2.0

; SEQ ID NO 1496  
; LENGTH: 578  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-925-300-1496

Query Match 58.4%; Score 2918; DB 10; Length 578;  
Best Local Similarity 99.1%; Pred. No. 7.3e-178;  
Matches 564; Conservative: 0; Mismatches 5; Indels 0; Gaps 0;

QY 395 DHQIVSVITEKALNAGFIQATDYVEITWQAYLDYLRVRVDFKQDSSKELEELRAAFTRALE 454  
Db 10 DESPTPTVTEKALNAGFIQATDYVEITWQAYLDYLRVRVDFKQDSSKELEELRAAFTRALE 69  
QY 455 YLKQVEEERFNSGDPSCVIMONWARIEARLCNNMOKARELWDSIMTRGNAKYANNMWLEY 514  
Db 70 YLKQVEEERFNSGDPSCVIMONWARIEARLCNNMOKARELWDSIMTRGNAKYANNMWLEY 129  
QY 515 YNLERAHGTQHCRCALHRAVQCTSDYPEHVCEVLLTMERTEGSLEDWDIAVQKTETRLA 574  
Db 130 YNLERAHGTQHCRCALHRAVQCTSDYPEHVCEVLLTMERTEGSLEDWDIAVQKTETRLA 189  
QY 575 RVNEORMKAAEKEAALVQOEESKAEQKRRARAEEKALKKKKIRGPEKRGADDEDEKQW 634  
Db 190 RVNEORMKAAEKEAALVQOEESKAEQKRRARAEEKALKKKKIRGPEKRGADDEDEKQW 249  
QY 635 DDEEOPSKRRRVENSIPAAGETQNVVEAAGPAGCAAVDVPEPPSKQKEKAASLRDMPK 694  
Db 250 DDEEOPSKRRRVENSIPAAGETQNVVEAAGPAGCAAVDVPEPPSKQKEKAASLRDMPK 309  
QY 695 VLHDSKSDSITVFSNLPYSMQEPDTKLRLPEACEGVVQIRPIFSNRGDFRGYCYVEFK 754  
Db 310 VLHDSKSDSITVFSNLPYSMQEPDTKLRLPEACEGVVQIRPIFSNRGDFRGYCYVEFK 369  
QY 755 EKKSALQALEMDRKSVEGRPMFVSPCVDKSKNPDKFVFRYSTSLEKHKLFISGLPFSCCTK 814  
Db 370 EKKSALQALEMDRKSVEGRPMFVSPCVDKSKNPDKFVFRYSTSLEKHKLFISGLPFSCCTK 429  
QY 815 EELEEICAHGTVKDLRLVTNRAGPKGLAYVEYENESQASQAVMKMDGWTIKENIIKVA 874  
Db 430 EELEEICAHGTVKDLRLVTNRAGPKGLAYVEYENESQASQAVMKMDGWTIKENIIKVA 489  
QY 875 ISNPPQKVPKPTRKAPGGPMLLPQTYGARGKGTQSLPRALQPSAAPOAENGPA 934  
Db 490 ISNPPQKVPKPTRKAPGGPMLLPQTYGARGKGTQSLPRALQPSAAPOAENGPA 549  
QY 935 AAAPAAVAPAAPEAPKMSNADFAKLF 963  
Db 550 AAAPAAVAPAAPEAPKMSNADFAKLF 578

RESULT 3

US-09-864-761-36026  
; Sequence 36026, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO  
; FILE REFERENCE: Aeo mica-x-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04

```

, PRIOR APPLICATION NUMBER: US 60/236,359
, PRIOR FILING DATE: 2000-09-27
, PRIOR APPLICATION NUMBER: PCT/US01/000666
, PRIOR FILING DATE: 2001-01-30
, PRIOR APPLICATION NUMBER: PCT/US01/000667
, PRIOR FILING DATE: 2001-01-30
, PRIOR APPLICATION NUMBER: PCT/US01/000664
, PRIOR FILING DATE: 2001-01-30
, PRIOR APPLICATION NUMBER: PCT/US01/000669
, PRIOR FILING DATE: 2001-01-30
, PRIOR APPLICATION NUMBER: PCT/US01/000665
, PRIOR FILING DATE: 2001-01-30
, PRIOR APPLICATION NUMBER: PCT/US01/000668
, PRIOR FILING DATE: 2001-01-30
, PRIOR APPLICATION NUMBER: PCT/US01/000653
, PRIOR FILING DATE: 2001-01-30
, PRIOR APPLICATION NUMBER: PCT/US01/000652
, PRIOR FILING DATE: 2001-01-30
, PRIOR APPLICATION NUMBER: PCT/US01/000661
, PRIOR FILING DATE: 2001-01-30
, PRIOR APPLICATION NUMBER: PCT/US01/000670
, PRIOR FILING DATE: 2001-01-30
, PRIOR APPLICATION NUMBER: US 60/234,687
, PRIOR FILING DATE: 2000-09-21
, PRIOR APPLICATION NUMBER: US 09/608,408
, PRIOR FILING DATE: 2000-06-30
, PRIOR APPLICATION NUMBER: US 09/774,203
, PRIOR FILING DATE: 2001-01-29
, NUMBER OF SEQ ID NOS: 49117
, SOFTWARE: Anomax Sequence Listing Engine
, SEQ ID NO 36026

```

```

/ OTHER INFORMATION: MAP TO AC008119.6
/ OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.1
/ OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.9
/ OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.8
/ OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.8
/ OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.2
/ OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.3
/ OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.6
/ OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.5
/ OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 2.5
/ OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.4
/ OTHER INFORMATION: EST_HUMAN HIT: A0124928.1, EVALUATE 7.000
/ OTHER INFORMATION: SWISSPROT HIT: Q98936, EVALUATE 1.80e+00
US-09-864-761-36026

```

[illegible]

```

: PRIOR APPLICATION NUMBER: PCT/US01/10542
: PRIOR FILING DATE: 2001-04-02
: PRIOR APPLICATION NUMBER: 60/236,384
: PRIOR FILING DATE: 2000-09-29
: PRIOR APPLICATION NUMBER: 60/194,118
: PRIOR FILING DATE: 2000-04-03
: NUMBER OF SEQ ID NOS: 27
: SOFTWARE: Patent In Ver. 2.0
: SEQ ID NO 27
: LENGTH: 687
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-969-384-27

Query Match
Best Local Similarity 21.2%; Pred. No. 1e-06;
Matches 155; Conservative 110; Mismatches 282; Indels 184; Gaps 36;

Qy 18 AGPKADGEEDVKAARTRRKVLGRAVAAPYKTMGPADQOQEGVSGDGEVAMASSAE 77
Db 2 AASTAAGKQIPKVAKVPKPAEVQITA-----EQLLRE-----AKERELE 43

Qy 78 SSQCEYEWYDEBEKKNQLEIER--LEEOLSIN-VYDYNCHVDLRLRLLEGELTKVPM 133
Db 44 LLPPPPQOKITDEELNDYKLKRKFTEDIRNKRNTVDSN-----WTKYAQWEESELEIOR 99

Qy 134 AR---QKMSIIFPLTELWLEWLDHETISMAQDGLDREHVVDLPFEKAVQYICPN-TWLEY 189
Db 100 ARSIYERALDQVDRNTLMLKYAEMEMKNFQ-----VNHARNIWDRAITTLPRVNOFWYKY 155

Qy 190 GOYSVVGIGQGGLEKVRSEFALSSVGLHMTKGLALWEAYREFESAIVEA-----241
Db 156 -TYMEEMLGNVAG--ARQVFER-----NMWQPEQAWHSYINFELRYKEVDRAITVE 206

Qy 242 -----ARLEKXHSIFRR-----QLAIPLDME-----ATFAEYENSE 274
Db 207 REVLVHPDVKNWIKYARFEKHAYFAHARKVYERAVEFFGDEHMDHLVYAFAKFENQK 266

Qy 275 DPICESVIQNKALQOLEKYKPYEALLQAEAPRLAEYQAYIDFEMKIGDPARIQILI--332
Db 267 EFERVRIYKY--ALDRISKDAQE-----LFKNYTIPEKKFGDRGIEDIIV 312

Qy 333 -----FERALVENCPLVDLWIRYSQYLDRLQVKDLVLSVHNRAIRNCP-----WT- 378
Db 313 SKRRFQYEEEVKANPHNYDAWFDYLRLVESDAEA-EAVREYVERAIANVPPIQEKRHWKR 371

Qy 379 -VALWSRYLLAMERHGYD---HQVTSVTEKALNAGFIQATDYV-----ETWQAYLDYL 428
Db 372 YIILWINYALYELEAKDPERTQVYQASLELIPHKKFTFAKMWILYAQFIRQKNLSLA 431

Qy 429 RRRVDF---KODSK-----ELEELRAAFTRALEYLKOVEERFNEGSDPSCVIMQNNWA 479
Db 432 RRALGTSGICPKPNKLFYVIELEQLUREDFRC-----RKLYEKFLFEGPENCSTWIKFA 486

Qy 480 RIEARLCNNMOKARELWDSITRGNAKYAN-MWLEYYNLERAHAGDQHCRCALHRAVQCT 538
Db 487 ELET-ILGDDIRARAIVELAISQPRLDMPVLKNSYIDFETEQEETERNLNRYRLRLQRT 545

Qy 539 SDYPEHVCEVLLTWER-----TEGSLEDWDIAVQKETTFLARVNRQRMK-----AAEKEA 588
Db 546 ----QHV-KVMYISFAQPELSGSGESL-----TKCROIYEANKTRMNCPEKEE 589

Qy 589 ALV-----QOEEKAPQRKRAAEK---KALKKKKKIRGPEKRGADDEDEKEW-----633
Db 590 RMLLESWRGFEESFGTASDKERVDKLMPKVKRRKVQ-----TDDGSDAGWEYFYD 643

Qy 634 --GDDEEQPS 642
Db 644 IFPEDAANQPN 654

RESULT 5
US-09-925-301-1012

```

Sequence 1012, Application US/09925301  
Patent No. US20020052308A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
FILE REFERENCE: PA106  
CURRENT APPLICATION NUMBER: US/09/925,301  
CURRENT FILING DATE: 2001-08-10  
PRIOR APPLICATION NUMBER: PCT/US00/05882  
PRIOR FILING DATE: 2000-03-08  
PRIOR APPLICATION NUMBER: 60/124,270  
PRIOR FILING DATE: 1999-03-12  
NUMBER OF SEQ ID NOS: 1694  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1012  
LENGTH: 708  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (18)  
OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (153)  
OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (229)  
OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (433)  
OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids  
US-09-925-301-1012

Query Match 4.5%; Score 225; DB 10; Length 708;  
Best Local Similarity 19.7%; Pred. No. 2.5e-06;  
Matches 145; Conservative 125; Mismatches 279; Indels 188; Gaps 34;

QY 14 AFSKAGKADGDEDEKAAARTKRV-----LSRAVAAATYKTMGPAMDQDEGV 63  
DB 19 AADMAASTAGQRIKPKYKAKVKNKAPAEVQITAEQLLREAKERELELLPP--PPQKID 76  
QY 64 ESDGDEYAMASSAESPGYEYDEEKNQLEI-----ERLEE-OLSNVVDY 112  
DB 77 EELNDYKLRKRK-----TFEDNIRKNTVISNMIKYAQWESLKEIQARSIVER 127  
QY 113 NCHVDL-----IRLLRLEGELTKVRMAR--QKXSEIFPLTEELWLEWHD 158  
DB 128 ALDVIDYRNTLWLYKAEMEMKRVQXAHARNIWDRAITTLPRVYNQFYKITYMEMLGNA 187  
QY 159 MAQDGLDRHVDLFEKAVKDYICPNWLEYQYQYVGGIGQKGLKRVSRPERALSSVG 218  
DB 188 GARQVFERMWEQPEQAWHSYI--NFEURYKE-----VDRARTIYER--XVL 231  
QY 219 LHMTGLALWEAYREFESAIVAARLEKVKHSLFRQLAIPLYDME-----ATAEYEE 271  
DB 232 VH--PDVKNWIKYAFEEKHAYFAHARKYI-----ERAVEFGDEHMDHEHLYVAFK 284  
QY 272 WSEDPIPEVIONKALQLEKYPYEEALLQAEAPRLAEOAYIDFEMKIGDPARQL 331  
DB 285 NQKEFERVVIKY--ALDRISKQDAQE-----LFKNYTFEKKFGDRRGIED 330  
QY 332 I-----FERALVENCIVLPDLWTRYQYLDRLQKVDLVLVSHNRAIRNC----- 376  
DB 331 IIVSKRRFOYEEVKANPHNYDAWFDYLRVSDAEAEAVREYVERAIANVPPIQEKRH 389  
QY 377 WT--VALMSRYLLAMERHGVHDQVLSVTFEKNALNAGFTQATDYVEIQWYLDYLR 434  
DB 390 WKRYIYLWNYALYEELEAKDPTRQVYQASLELIPHKFTFKXKWILYAFQFIRQ--- 446  
QY 435 KODSSKELEELRAAFTRAL-----EYLUKQVE-----ERFNESGDPSCV 473  
DB 447 -----KNLSLARRALGTSIGKPKNKLKFKVYIELEQLREDFRCRKLKFKLEFGP 501

QY 474 IMQNWARIEARLCNNMKARELWDSIMTRGNAYAN-MWLEYYNLERAHGDTQHCRAH 532  
DB 502 SWIKFAELET-ILGDIDRARIYELAISQRLDMPVLWKSVIDFEIQEETEETRNLYR 560  
QY 533 RAVQCTSDYPEHVCEVLLTMR-----TEGSLEWDIDIAVQKTETRLARVNQRMK---- 582  
DB 561 RLLQRT---QHV-KVWISFAQFELSSGKESL-----TKRQIYEANKTRN 604  
QY 583 AAEKEAALV-----QDEEKAQKRAERK---KALKKKKIRGPKRGADDEDEKEM 633  
DB 605 CEEKERLMLLESWRSFEEFCTASDKERVDKLMPEKVKRRKVQ-----TDDGSDAGW 658  
QY 634 -----GDDEEQPS 642  
DB 659 BEYFDYIRPEDAANQPN 675

## RESULT 6

US-09-978-242-3  
Sequence 3, Application US/09978242  
Patent No. US20020098566A1  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
Yue, Henry  
Corley, Neil C.  
Shah, Purvi  
TITLE OF INVENTION: HUMAN NUCLEOLIN-LIKE PROTEIN  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.,  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/978,242  
FILING DATE: 15-Oct-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/241,333  
FILING DATE: <Unknown>  
APPLICATION NUMBER: 08/990,114  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0451 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-855-0555  
TELEFAX: 650-845-4166  
TELEX: <Unknown>  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 714 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 128842  
SEQUENCE DESCRIPTION: SEQ ID NO: 3:  
US-09-978-242-3

Query Match 4.5%; Score 224.5; DB 10; Length 714;  
Best Local Similarity 23.5%; Pred. No. 2.7e-06;  
Matches 94; Conservative 67; Mismatches 134; Indels 105; Gaps 16;



QY	583	AAEKTAALVQOEBAEKAOQRARAB-----KKALKKKKKIRGEKKGAEDDE	630
Dd	187	ASEDEDEDEEEDEEEDDSEEAEMBITPAKKGAKAPKVVPVPAKNVAEDDDDD	246
QY	631	KWGGDDEEQ-----PSKRRR---VENSIPAAGTQNV---	661
Dd	247	EDEDDEDEDEDEEEEEEPVKPAGRKKEMTQKEVPEA-KKQVEGSEST	305
QY	662	-----VAAGPAGK--CAAVDVPPSPSK-	686
Dd	306	TPENLFIGNLNPNKSVAELKVAISEPPAKNDLAAVDVRTCTNRKFVGVDFAEDLEKAL	365
QY	687	SL-----KRDPKVLHDSK--DSITVFVSNNLPYSMQBPDTKLRPLFPACGEVVQ	734
Dd	366	EITGLKVFGNETIKLEPKPG-RDSKKVRAARTLLAKNISFNITIDE--LKFEVF--DALE	419
QY	735	IRPIFNSNGDFRGYCYVEFKKEKSAQALEMDR-KSVEGRPMFVSPCVDKSNPDFKVR	793
Dd	420	IR-LVSDQGKSGIAYIEFKSEADAKNLEBKQAEIDGRSVLSYYTGEGQQOETGNK	478
QY	794	YSTLESKHKLFTISGLPFSCTEEELEEICKAHGTVKDLRLVTNRAGPKGLIAYVEYESNQ	853
Dd	479	STWSGESKTVLVLSNLSYSATEETLQEVFERATFIK---VPQNQOGKSGVAFTEFASF	535
QY	854	ASQAVMKDGMTIKENIIKVAISNPPOKPYKPETRPAP	893
Dd	536	AREALNSCNMBIEGTRTLRELQGP-----RGSPNARSOP	570

## RESULT 7

```

US-09-864-761-34622
; Sequence 34622, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aemica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30

```

```

; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annamux Sequence Listing Engine vers. 1.1
; SEQ ID NO 34622
; LENGTH: 41
; TYPE: PR1
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC008119.6
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.9
; OTHER INFORMATION: SWISSPROT HIT: Q32866, EVALU6 6.10e+00
; OTHER INFORMATION: EST_HUMAN HIT: AU143744.1, EVALU6 1.00e-11
; US-09-864-761-34622

```

```

Query Match      4.3%  Score 217;  DB 10;  Length 41;
Best Local Similarity 100.0%;  Pred. No. 2.2e-07;
Matches 41;  Conservative 0;  Mismatches 0;  Indels

```

## RESULTS

```

RESUL1 8
; Sequence 164, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Glatt, Karen
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Hoersh, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; TITLE OF INVENTION: OF CERVICAL CANCER
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 164
; LENGTH: 1938
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-171-311-164

```

Query Match	4.2%	Score 208.5;	DB 9;	Length 1938;
Best Local Similarity	19.6%	Pred. No. 0.0001;		
Matches 213:	Conservative 170;	Mismatches 399;	Indels 305;	Gaps 44;



Db 1485 AREKTRALSARALEEAEKEELERTNKMKAEMEDLVSKDVGKNVHELEKSKRAL 1544  
Qy 550 LT-MERTEGSLDWDIAVOKTETRLAR--VNEORMKA-----AKEAALVOQ 593  
Db 1545 ETQMEEMKTOLEEELEDELOQATEDAKLRLEVNQALKGQFERDLOARDEQNEEKRRQLOQ 1604  
Qy 594 -----EEKABQQRKRAAEK-----ALKKKKKIRGPEKRG 624  
Db 1605 LHEYTELEDERKQALAAAKKLEGLDKDLELQADSAIKGREAIKQLRKLQAKMDF 1664  
Qy 625 ADEDEKEWGDDE-----EEQPSKRRRVENSIPAAGETQNVVAAGPAGCAAVDVEPPS 679  
Db 1665 QRELEDRASRDEIFATAKENEKAKSLEADLMOLQE-----DLAAERARKQA-DLEKEE 1719  
Qy 680 KOKEKAASIKRDMPKVLHDSKDSITVFSNLPYSMOEPDTPKLRPLFEACGEVQIRPIF 739  
Db 1720 LAELASLSG--RNAQD-EKRRLEARIAQLEEEEEEQGNMAMSDRVRKATQQAOL 1776  
Qy 740 SNRGDFRGYCYVEFEKESALQALEMDRKSVEGRPMFVSPCVDKSKNPDPKVFRYSTSLE 799  
Db 1777 SN-----ELATERSTAQKNESARQOLERQ-----1800  
Qy 800 KHKLFISGLPSCCTKEELEICKAHGTVKDLRLVTNRAGKPKGLAYVEYE-----NESQA 854  
Db 1801 -----NKLRSKLHEMEGAVKSKFKSTIAALEAK-IAQLEEQVEQEAAREKQA 1846  
Qy 855 SOAVMKMDGMTIKENIKVAISNPPORVKPEKPTRKAPGGPMLLPQTYGARGKRTOLS 914  
Db 1847 ATKSLKQDKKLIKLEILQV-----EDERKMAEQYKEQAE-----KGNARVK 1887  
Qy 915 LLPRALQ 921  
Db 1888 QLKRLQE 1894

## RESULT 10

US-10-171-311-162  
; Sequence 162, Application US/10171311  
; Publication No. US20030087270A1  
; GENERAL INFORMATION:  
; APPLICANT: Schlegel, Robert  
; APPLICANT: Chen, Yan  
; APPLICANT: Zhao, Xumei  
; APPLICANT: Monahan, John  
; APPLICANT: Kamatkar, Shubhangi  
; APPLICANT: Glatt, Karen  
; APPLICANT: Gannavarapu, Manjula  
; APPLICANT: Hoerish, Sebastian  
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR  
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY  
; TITLE OF INVENTION: OF CERVICAL CANCER  
; FILE REFERENCE: MRI-035  
; CURRENT APPLICATION NUMBER: US/10/171,311  
; CURRENT FILING DATE: 2002-06-12  
; PRIOR APPLICATION NUMBER: US 60/298,159  
; PRIOR FILING DATE: 2001-06-13  
; PRIOR APPLICATION NUMBER: US 60/298,155  
; PRIOR FILING DATE: 2001-06-13  
; PRIOR APPLICATION NUMBER: US 60/335,936  
; PRIOR FILING DATE: 2001-11-14  
; NUMBER OF SEQ ID NOS: 238  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 162  
; LENGTH: 1972  
; TYPE: PRF  
; ORGANISM: Homo sapiens  
US-10-171-311-162

Query Match 4.2%; Score 208.5; DB 9; Length 1972;  
Best Local Similarity 19.6%; Pred. No. 0.0001;  
Matches 213; Conservative 170; Mismatches 399; Indels 305; Gaps 44;

Qy 6 ETSASEPEAESK--AGPKRADGEE--DEVKAAARTRRKVLRSRAVAAYATKTMGP---AWDQO 58  
Db 935 EDRGQLOAERKKAQMOQLDLEEQLEEEEAARQKIQ-LEKVTAAEKIKKLEDEILVMDQ 993  
Qy 59 -----BEGVSE-----SOGDEYA-----MASSAESPGYEWYEDDEEKNQLE 97  
Db 994 NNKLSKERKLLLEERISDUTTNLAEEERAKNLTJLKNKHESMISELYRL-KKEEKSQOE 1052  
Qy 98 IERLEEQLSINVDYNCHVDLIRLRLRGELTKVEMARQKMEIPLPTEELWLEWHDEI 157  
Db 1053 LEKLRKLEGASDFH-----EQIADLQQAELAKMLAKKEEL-----QAALRLDDEI 1103  
Qy 158 SMAQDGLDR-----EHVYDLFEKAVKDYICPNWILEYQYSGVGGIGKGLKVRSVFER 212  
Db 1104 AQKNNAKKIRELEGHISDLQEDLDSARAARN---KAEKOKRDLGEE--LEALTELED 1157  
Qy 213 ALSSVGLHMTKGLALWEAYREFESAIVEAARLEKVSLSFRRLQATPLPYDMEATFAYEEM 272  
Db 1158 TLDSTATQOE-----LRAKREQEVTVLKALDEETRS-----HEAQVQEMROK 1200  
Qy 273 SEDPIESVIQ-----NYNKALQOLEKYKPYEEALLQAEAPRLAEYQAYIDPEMK- 322  
Db 1201 HAQAVEELTEQEQFKRAKANLDKNNKQTLK-----ENADLAGELVLQAGKEVEHKKK 1256  
Qy 323 -----IGDPARIQIIFERALVEN-----CLVPDLWIRYSQ 352  
Db 1257 LEAQVOELQSKSDGERARAEALNDKVHLQNEVESVTGMLNEAECKAIKLAQDVASLSQ 1316  
Qy 353 YLDRLQKVKD-----LVLSVHNRAIRNCPWTVALWSRYLLAMERHGVHDHV-----ISVT 402  
Db 1317 LQDTQELQOETROKLVNSTKLRLQLEE-----ERNSLQDQDDEMEAKON 1361  
Qy 403 FEKALNAGFIQATDYVEITWOAVLDYLRVRVDFKODSSKELELRAAF-TRALEYLK-QEV 460  
Db 1362 LERHISTUNIQLSDSKKLLQDFASTVEALEBEGKKRFQKEIENLTQYBEKAAAYDKLEKT 1421  
Qy 461 EERFNEGDPSCVIMONWARIEARLCNNMKARELWDSIMT-----RGNAK 506  
Db 1422 KNRLQOELDLVLDLN---QRLVSNLEKKORKFDOLLAEBEKNISSYADDERARAE 1477  
Qy 507 YANMWLEYNLERAHGDTQHCRAKHRA-----VQCTSDYPEHVCEV-----L 549  
Db 1478 AREKETKALSARALEEAELEKEELERTNKMKAEMEDLVSKDVGKNVHELEKSKRAL 1537  
Qy 550 LT-MERTEGSLDWDIAVOKTETRLAR--VNEORMKA-----AKEAALVOQ 593  
Db 1538 ETQMEEMKTOLEEELEDELOQATEDAKLRLEVNQALKGQFERDLOARDEQNEEKRRQLOQ 1597  
Qy 594 -----EEKAEQQRKRAAEK-----ALKKKKKIRGPEKRG 624  
Db 1598 LHEYTELEDERKQALAAAKKLEGLDKDLELQADSAIKGREAIKQLRKLQAKMDF 1657  
Qy 625 ADEDEKEWGDDE-----EEQPSKRRRVENSIPAAGETQNVVAAGPAGCAAVDVEPPS 679  
Db 1658 QRELEDRASRDEIFATAKENEKAKSLEADLMOLQE-----DLAAERARKQA-DLEKEE 1712  
Qy 680 KOKEKAASIKRDMPKVLHDSKDSITVFSNLPYSMOEPDTPKLRPLFEACGEVQIRPIF 739  
Db 1713 LAELASLSG--RNAQD-EKRRLEARIAQLEEEEEEQGNMAMSDRVRKATQQAOL 1769  
Qy 740 SNRGDFRGYCYVEFEKESALQALEMDRKSVEGRPMFVSPCVDKSKNPDPKVFRYSTSLE 799  
Db 1770 SN-----ELATERSTAQKNESARQOLERQ-----1793  
Qy 800 KHKLFISGLPSCCTKEELEICKAHGTVKDLRLVTNRAGKPKGLAYVEYE-----NESQA 854  
Db 1794 -----NKLRSKLHEMEGAVKSKFKSTIAALEAK-IAQLEEQVEQEAAREKQA 1839  
Qy 855 SOAVMKMDGMTIKENIKVAISNPPORVKPEKPTRKAPGGPMLLPQTYGARGKRTOLS 914  
Db 1840 ATKSLKQDKKLIKLEILQV-----EDERKMAEQYKEQAE-----KGNARVK 1880  
Qy 915 LLPRALQ 921







```
QY 542 PEHV-----CEVLLTMEGTSGLEDWDIAVQKT 569
:|
Db 546 -QHVKVWISFAQFELSSRRKRRKFAKRCROIYEKLTKPCE---TVKKRRDLCCWN----- 595
QY 570 ETRLARVNEQRMKAKEKAALVQQEEEEKAEQORRRAR----- 605
Db 596 ---LGEVLEEEFEGTASDKERYDKLMPKVKRRKRYQTDGSDAGWEEYFDYIFPEDAANQ 652
QY 606 -----AEKKALKKKKIRGPEKRGADED-DEKE 632
Db 653 PNKLLAMAKLWKQOQKEGDAEHHPDEDVDESE 686
```

Search completed: June 18, 2003, 14:01:09  
Job time : 60 secs

**THIS PAGE BLANK (USPTO)**









F:555-619/Domain: ribonucleoprotein repeat homology <RRM4>

Query Match 5.9%; Score 293.5; DB 2; Length 705;  
Best Local Similarity 23.1%; Pred. No. 3.5e-08;  
Matches 103; Conservative 71; Mismatches 147; Indels 125; Gaps 13;  
QY 556 EGSLEWDIAVQKTRTLARV  
DB 37 ESSEEMEVPVKTPAKKAATPAKATPGKAATPAKATVTPGKKGATPGAKNGKQAKQES 96  
QY 587 EAALVQOEKAFORRAAEKALKKKKIRGPEKRGADDE 630  
DB 97 EEEEDSDDEQDKPIKPIAKKAVAKKESEEDDEDESEEEKPVAKRPLAKNPPVK 156  
QY 631 -----KWDDEEOPSKRRRVENSIPA--AGETQN-----VEVAA 664  
DB 157 VTPAKSTGKQESDEESEEKPVAKPKAKRPAKRGKQSEDESEDEEPEPEVAS 216  
QY 665 GPAGKCAAV-----DVEPPSKQKEKAASLRDMP-----KYL 696  
DB 217 AOKGKTAPAKAEDDDDDDDDDDEDEDDDDQCGPAKRKKEMPKNNVPEAKTK 276  
QY 697 HDSSKDSIVFVSNLPYSQEPDTK--LRPLFEACGEVVOIRPISNRRGDFRCYCYVEK 754  
DB 277 TDTASEGFSIFIGNLSTLDFDLKDLREFTSKKNLTIDVRI-----GSKKFGYVDS 332  
QY 755 EKSALQALEMDRKSVEGRPMFVSPCV--DKSKNPDFKVFYRSTSLSKHLFISGLPFS 812  
DB 333 SEEEVEKALKSKKILGLVEIAEKAVAFDKNAENKKEK-----DSRTLFVNIPYST 387  
QY 813 TKEELEICAKGTWKDLRLVNRACKPKGLAYVEYENESQASQAVMKMDGMTIKENI 872  
DB 388 SAEELQEIPE--NAKDIRIPTGNDGSKNGIAYVEFSTEAEANKALEEKGAEIGRS 444  
QY 873 VALSNPPQRKPEKPT--RKAPCG 895  
DB 445 VDFTG-----EKSONSGRRGPAG 463

## RESULT 6

T31808  
hypothetical protein M03F8.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 15-Sep-2000  
C:Accession: T31808  
R:Davidson, S.; Wohldmann, P.  
submitted to the EMBL Data Library, July 1997  
A:Description: The sequence of C. elegans cosmid M03F8.  
A:Reference number: 221089  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-761 <D>  
A:Cross-references: EMBL:AF016441; PIDN:AAB65909.1; GSPDB:GN00023; CESP:M03F8.3  
A:Experimental source: strain Bristol N2; clone M03F8  
C:Genetics:  
A:Gene: CESP:M03F8.3  
A:Map position: 5  
A:Introns: 43/3; 388/3; 459/3; 632/2  
C:Superfamily: Schizosaccharomyces pombe hypothetical protein SPBC31F10.11c

Query Match 5.5%; Score 276.5; DB 2; Length 761;  
Best Local Similarity 19.9%; Pred. No. 3.1e-07;  
Matches 152; Conservative 132; Mismatches 268; Indels 211; Gaps 31;  
QY 30 KAARTRKV-----LSRAVAATYKTGPAWQ--QEEGVSESDGDEYAMASAESPGY 83  
DB 19 KAAKVNKYVSFADFINSKVNITY-FIAPAQLQITAEQLLREAKERELELIPAPKTKITD 77  
QY 84 EWEYDEEEKNOLETERLEEQLSINVDYNCHVDLIRLLRLP-----GELT 129  
DB 78 PDELKEYQRRKKEFE-----DGIRNRMLANWIKYKWEESIGEIQ 120

QY 130 KVRMARQKMEIFPLTEELWLEWLEHDEISMAQDGLDREHYVDFEKAVKDYICP-----NIW 186  
DB 121 RARSFERALDVEDHRSISISIWQYAEEMRCKQ-----INHARNVDFRAIT--IMPRAMQFW 174  
QY 187 LEYGOYSGVGICQKGGLEKV-----RSVFERALSGLHMTKGLALWEAYREPESATV 239  
DB 175 LKYSY-----MEEVINIFGARQIFER-----WIEWEPPEQAMQTYINFELRYK 218  
QY 240 EAARLEKVHSLFRROLAIPLYDME--ATPAEYEE-----WSEDDPIPE 279  
DB 219 E--IDRAVSYORFLVHGIVNQWIKYAKEEERNGYIGNARAAYEKAMEYFGEDINE 275  
QY 280 SVIQNYNKALQOLEKYPYEEA-----LLOAEAPRLAE--YQAYIDFEMKIGDPARIOL 331  
DB 276 TVLVAF--ALPE-ERQKEHERARGIFKYGLDNLPSNRTEEIEFKHYTHQHEKKFGERVIG 332  
QY 332 I-----FERALVENCLVPLWIRYSQYLDORQLKVDLVSVHNRAIRNCP--WTVAL 381  
DB 333 VIISKRKTOYKEMVSENGYNDADFYLRLLENEETDREEDVYVERAIAINIPPHSEKRY 392  
QY 382 WSRYLLAMERHGVHDHVISVTFEKAALNAGFIQATD-----YVEIQAYLDYLRRRYD 433  
DB 393 WRYYIYLVINLYALYELVAKDFDRARQV--YKACIDIIPHKTTFKAVWIMFAHFETRLD 451  
QY 434 F-----KODSSK-----ELEELRAAFTRALEYLKQVEERFNESEGDPSCVIM 475  
DB 452 LNAARKINGVAIGKCPKDKLFRAYIDLEQLREFDRC-----RKYKEKLESSPESQTM 506  
QY 476 QNWARIEARL-----CENNQKARELWDSIM 500  
DB 507 IKFAELETLLGDTDRSAVFTTIAVQPALDMPPELLWKAYIDFEIACEEHEKARDLYETLL 566  
QY 501 TRGNKAYANMWLEYYNLERAHGDTQHCRAKALHRAVQCTSDYPEH-----VCEYLLTWERT 556  
DB 567 QRTN--HKVWISMAEFEGTICNFEGARKAFERANQSLNENAEKEERLMLELAWKECETKS 624  
QY 557 GSLEDWDIAVQKTETRLARVNEQRMAKAEKAAALVQOE-----BK 597  
DB 625 GDQE-----ALKRVETMPRRVKRKQIQETDGVDAWEEYFDYIFPDQAAKGSFKLLEA 680  
QY 598 AEQRKRAAEKKALKKKKKIRGPEKRGADDEDEKWDDEDEEQ 640  
DB 681 AARWKRER--EAAARAAQELDAPIPGDDDEEKEAGKDAEAK 722

## RESULT 7

A39205

nuclear localization sequence-binding protein NSR1 - yeast (Saccharomyces cerevisiae)  
N:Alternate names: nucleolar protein NSR1; p67 protein; protein G7001; protein YGR159  
C:Species: Saccharomyces cerevisiae  
C:Date: 14-Feb-1992 #sequence\_revision 14-Feb-1992 #text\_change 24-Sep-1999  
C:Accession: A39205; PC2395; S60450; S64468; S64470  
R:Lee, W.C.; Xue, Z.; Melese, T.  
J. Cell Biol. 113, 1-12, 1991  
A:Title: The NSR1 gene encodes a protein that specifically binds nuclear localization  
A:Reference number: A39205; MUID:91177946; PMID:1706724  
A:Accession: A39205  
A:Molecule type: DNA  
A:Residues: 1-414 <LE>  
A:Cross-references: GB:X57185; NID:g4057; PIDN:CAA40472.1; PID:g4058  
R:Gamberi, C.; Contreas, G.; Romanelli, M.G.; Morandi, C.  
Gene 148, 59-66, 1994  
A:Title: Analysis of the yeast Nsr1 gene and protein domain comparison between Nsr1 a  
A:Reference number: PC2395; MUID:95011659; PMID:7926838  
A:Accession: PC2395  
A:Molecule type: mRNA  
A:Residues: 155-414 <GAM>  
A:Cross-references: EMBL:X57185  
R:Note: The authors translated the codon TTC for residue 188 as Arg, GGT for residue  
R:Skala, J.; Nawrocki, A.; Goffeau, A.  
Yeast 11, 1421-1427, 1995  
A:Title: The sequence of a 27 kb segment on the right arm of chromosome VII from Sacc





C:Superfamily: nucleolin; ribonucleoprotein repeat homology  
C:Keywords: DNA binding; nucleolus; nucleus; phosphoprotein; RNA binding  
F:308-373/Domain: ribonucleoprotein repeat homology <RRM1>  
F:394-456/Domain: ribonucleoprotein repeat homology <RRM2>  
F:487-550/Domain: ribonucleoprotein repeat homology <RRM3>  
F:573-634/Domain: ribonucleoprotein repeat homology <RRM4>

Query Match 4.9%; Score 245.5; DB 2; Length 707;  
Best Local Similarity 24.2%; Pred. No. 1.3e-05;  
Matches 92; Conservative 64; Mismatches 127; Indels 97; Gaps 14;

QY 595 EEKAEQKRAAEKKA-----LKKKKKIRGPKRGADE-----DEKEWGD 635  
DB 207 EEEAMETTPAKGKAAKVVVKAKNVAEDEDEDEDEDEDEDEDEDEDEDEDE 266  
QY 636 DEEQP-----SKRRRENSIP----- 652  
DB 267 EEEPEKAEAPGKRRKKEMAKQAPEAKQKVEGTEPTAFNLVGNLNFNKSAPELKTG 326  
QY 653 -----AAGETONVEAAGPAGCAADVDEPPSKQKKAASL-----KRDMPKVLH 697  
DB 327 ISDVFAKNDLAVDVIGTRKRGVDFE-SAEDELEKALELTGLKVFGNEIKLEKPG-K 384  
QY 698 DSSK--DSITVFNLSNLSYMOEPTKRLPLFEACEGVVQIRPISNRGDFRGYCYVEPK 755  
DB 385 DSKKERDARTLLAKNLPYKVTQDE--LKEVFEAAEI-----RLVSKDGKSGKIAYIEPKT 438  
QY 756 EKSALOALEMDR-KSVEGRPMFVSPCVDKSNPKDFKFRYST-SLEKHKLFISGLPESCT 813  
DB 439 EADAEKFEKQGTEDRSLSLYTGTGKQNDYRGKNTWSGESKTLVLSNLSYSAT 498  
QY 814 KEELEEICAKHTGKVDRLVTRNRAKPKGLAYVEYENESQASQAVMKMDGMTIKENIYKV 873  
DB 499 EETLQEVFEKATFIK---VPQNGKSGKYAFIEFASFEDAKEALNSCNKEIEGRAIRL 555  
QY 874 AISNPPQKRVPEKPTRKAP 893  
DB 556 ELQGP-----RGSPNARSQP 570

RESULT 11  
DNMS  
nucleolin - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 30-Sep-1992 #sequence\_revision 30-Sep-1992 #text\_change 22-Jun-1999  
C:Accession: A29958; A40769; A56240; I84688  
R:Bourbon, H.M.; Lapeyre, B.; Anallie, F.  
J. Mol. Biol. 200, 627-638, 1986  
A:Title: Structure of the mouse nucleolin gene. The complete sequence reveals that each  
A:Reference number: A29958; MUID:88316930; PMID:3137346  
A:Accession: A29958  
A:Molecule type: DNA  
A:Residues: 1-707 <BOU>  
A:Cross-references: GB:X07699; NID:g53453; PIDN:CAA30538.1; PID:g53454  
R:Pasternack, M.S.; Bleier, K.J.; McInerney, T.N.  
J. Biol. Chem. 266, 14703-14708, 1991  
A:Title: Granzyme A binding to target cell proteins. Granzyme A binds to and cleaves nucleolin  
A:Reference number: A40769; MUID:91317840; PMID:1860869  
A:Accession: A40769  
A:Molecule type: protein  
A:Residues: 2-20, 'X', '22-24 <PAS>  
R:Yang, T.H.; Tsai, W.H.; Lee, Y.M.; Lei, H.Y.; Lai, M.Y.; Chen, D.S.; Yeh, N.H.; Lee, S.  
Mol. Cell. Biol. 14, 6068-6074, 1994  
A:Title: Purification and characterization of nucleolin and its identification as a transmembrane protein  
A:Reference number: A56240; MUID:94344117; PMID:8065340  
A:Accession: A56240  
A:Molecule type: protein  
A:Residues: 2-19; 558-567 <YAN>  
R:Bourbon, H.  
Gene 68, 73-84, 1988  
A:Title: Sequence and structure of the nucleolin promoter in rodents: Characterization of the promoter  
A:Reference number: I48118; MUID:89121496; PMID:2906027  
A:Accession: I84688

A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-44 <RES>  
A:Cross-references: GB:M22089; NID:g200111; PIDN:AAA39841.1; PID:g554246  
C:Comment: This housekeeping protein is involved in the synthesis, packaging, and maturation of ribosomes.  
C:Genetics:  
A:introns: 6/3; 45/3; 211/1; 273/1; 302/1; 349/2; 391/1; 432/3; 484/1; 525/2; 566/1;  
C:Superfamily: nucleolin; ribonucleoprotein repeat homology  
C:Keywords: DNA binding; duplication; nucleus; phosphoprotein; RNA binding; transcript processing  
F:310-375/Domain: ribonucleoprotein repeat homology <RRM1>  
F:311-316/Region: RNA-binding RNP2 motif  
F:349-356/Region: RNA-binding RNP1 motif  
F:396-458/Domain: ribonucleoprotein repeat homology <RRM2>  
F:397-402/Region: RNA-binding RNP2 motif  
F:431-438/Region: RNA-binding RNP1 motif  
F:488-551/Domain: ribonucleoprotein repeat homology <RRM3>  
F:489-494/Region: RNA-binding RNP2 motif  
F:524-531/Region: RNA-binding RNP1 motif  
F:570-634/Domain: ribonucleoprotein repeat homology <RRM4>  
F:571-576/Region: RNA-binding RNP2 motif  
F:607-614/Region: RNA-binding RNP1 motif

Query Match 4.9%; Score 244.5; DB 1; Length 707;  
Best Local Similarity 24.7%; Pred. No. 1.4e-05;  
Matches 99; Conservative 65; Mismatches 138; Indels 99; Gaps 15;

QY 553 ERTEGSLDWDIAVQKTETRLARVNEQRMKAEEALVQOEKAEQKRAAEKKALK 612  
DB 201 EDDDEEDDEEVEEITTAAGKKTAKVVPMAKSAVEEDDEDE----- 248  
QY 613 KKKIRGPKGADEDEKWDGDD-----EEEP-----SKRRRV 647  
DB 249 -----DEDEDEDEDEDEDEDEDEDEEPEEYKAAAGKRRKKEWTKQEAPEAKQKV 300  
QY 648 ENSIP-----AAGETONVEAAGPAGCAADVDEPPSKQKKAASL----- 679  
DB 301 EGSEPTTFNLFIGNLPKNSYNELKFAISLFAKNDLAVDVTRTKRGYVDFE-SA 359  
QY 680 KOKEKAASL-----KRDMPKVLHDSKK--DSITVFNLSNLSYMOEPTKRLPLFE 727  
DB 360 EDLEKALELTGLKVFGNEIKLEKPG-KRDSKVVRAARTLLAKNLSFNTEDE--LKEVFE 416  
QY 728 ACGEVVOIRPISNRGDFRGYCYVEFEKESALQALEMDR-KSVEGRPMFVSPCVDKSN 786  
DB 417 --DAMEIR-LVSQDGKSGKIAYIEFSEADAENLEKQGAIEDGRSVLSYTTGKQGR 472  
QY 787 PDFKVFYRST-SLEKHKLFISGLPSCYKBELEECATGTVKDLRLVTRNRAKPKGLAY 845  
DB 473 QE-RTGKTSTWSGESKTLVLSNLSYSATKETLEEYFEKATFIK---VPQNGKSGKYAF 528  
QY 846 VEYENESQASQAVMKMDGMTIKENIYKVAISNPPQKRVPEK 886  
DB 529 IEFASFEDAKEALNSCNKMEIGRTIRLELOGSNSRSQPSK 569

RESULT 12  
A96527  
probable num1 protein [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: A96527  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon  
Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar,  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzborg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.;  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: A96527

A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-557 <STO>  
A:Cross-references: GB:AE005173; NID:g11094815; PIDN:AAG29744.1; GSPDB:GN00141  
C:Genetics:  
A:Gene: F27K7.6  
A:Map position: 1

Query Match 4.9%; Score 243; DB 2; Length 557;  
Best Local Similarity 24.1%; Pred. No. 1.3e-05;  
Matches 88; Conservative 71; Mismatches 132; Indels 74; Gaps 16;

QY 561 DMDIAVQKTRLARVNEORMKAAKEALVQOEKEAFQKRAAEKALKKKKIRGP 620  
DB 192 DEDSEDEKATKAA--PAAAKAASSSDSDSDSDSESEDEK--PAQKADTKASKSS 247  
QY 621 ERGADEDEKEWGDDEEQPSKRRRVENSIPAAGETQNVVAAGPAGCAVDVPPSK 680  
DB 248 DSSSESEDESE--DEETPKKK-----SSDVEM-----VDAEKSSA 282  
QY 681 OKKAASLRDMPKVLHDSKSDSITVFSNLPYSMQEPTDKLRPLFEACGEVVOIRPFS 740  
DB 283 KQPKTPST-----PAAGSKTLFAANLSFNIERAD--VENFFKEAGEVVDVR--FST 330  
QY 741 NR--GDFRGYCYVFEKEKSALQALEMDRKSVEGRPMFVSPCVDKSKNPDFKVFYRSTSL 798  
DB 331 NRDDGSRFGHVEFASSEAKALEFHGRPLLRGRIIDIAQERGERPAFTPOS 390  
QY 799 -----EKHLFTISGLPFCSTKEELEIEICAH-----GTVKDLRLVTR--AGKPKGLAY 845  
DB 391 FRSGDGGDEKKIFVKGDFDASLEDDIKNTLREHFFSGGEIKNVSVPIDRTGNSKGIAY 450  
QY 846 VEYENESQASQAVKMDGMTIKENIKVAISNPPQKVPKPKAPGPGMMLLPQTYCA 905  
DB 451 LEF---SEGKEKALELNGSDMGGGFYL--VDEP-----RPRGDSGGGGF-----G 492  
QY 906 RKGR 910  
DB 493 RGNR 497

## RESULT 13

T17297

hypothetical protein DKF2p586f1023.1 - human

C:Species: Homo sapiens (man)

C&gt;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 31-Jan-2000

C:Accession: T17297

R:Ansorge, W.; Wirkner, U.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.

submitted to the Protein Sequence Database, September 1999

A:Reference number: 218729

A:Accession: T17297

A&gt;Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-960 &lt;ANS&gt;

A:Cross-references: EMBL:AL117547

A:Experimental source: adult uterus; clone DKF2p586f1023

C:Genetics:

A:Note: DKF2p586f1023.1

C:Superfamily: unassigned ribonucleoprotein repeat-containing proteins; ribonucleoprotein

Query Match 4.8%; Score 239.5; DB 2; Length 960;  
Best Local Similarity 22.4%; Pred. No. 4e-05;  
Matches 116; Conservative 80; Mismatches 173; Indels 149; Gaps 22;

QY 404 EKALNAGTQATDY--VEITQAYLDYLRVRVDFK-----QDSSK---ELEELA 447  
DB 59 QKHFNKSIFDTRITVECKSEGDPAKPRANSKHAQPSQPKPPKDSSTTPEIKKDEKK 118  
QY 448 AFTRALEYLKQVEERFNESGDPSCVIMQNWARIEARLNCNNQAKRELWDSITMRGNAKY 507  
DB 119 KVAGOLEKLKEDTE-----FOEFLSVHORRAQAATWANDGLDAEPSKSKP 165  
QY 508 ANMWLEYNLEAAGDTHCRKALHRAVQCTSDIPEHVCEVLLTWERTSGSLEDWDIAVQ 567

DB 166 AS---DYLNFSDSQE-----SEEGAGED----- 188  
QY 568 KTETRLARVNEORMKAAKEALVQOEKEAFQKRAAEKALKKKKIRGPEKGADE 627  
DB 189 -----LEEASL-----EPKAAVQKEL--SDMDYLSKMKVWKGAGSSSSSEEE 227  
QY 628 DDEKE-----WGDDDEEQPSKRRRV--ENSIPAAGETQNVVAAGPAGKC-----AAVDVE 676  
DB 228 ESEDAVHCDGSEAEFEEDSSATPVQERDSKGAGEQGM-----PACKKKRPPPEARATE 282  
QY 677 PPSKQKEKAAALSRDMPKVLHDSKSDSITVFSNLPYSMQEPTDK--LRPLFEACGEVVO 734  
DB 283 KPAQKE-----PTTCH-----TVKLRGAPFNVNTEKNVMEFLAPL-----KPVA 321  
QY 735 IRPFSNRGDFRGYCYVFEKEKSALQALEMDRKSVEGRPMFVSPCVDKSKN--PDFKVFR 793  
DB 322 IRIVRNAHGNKTYGIFYDFVSNEEVQALKCNREYMGGR--YIE--VFREKNVPTTKGAP 377  
QY 794 YSTS-----LEKHKLFTISGLPFCSTKEELEIEICAHGTVKDLRL--VTN 835  
DB 378 KNTTSMOGRILGENEEEDLAESGRLFVRNLPYTTSTEEDLEKLFSGYGLSELHVPIDS 437  
QY 836 RAGRPKGLAYVEYENESQASQAVKMDGMTIKENIKV 873  
DB 438 LTKPKGFAFITFPEHVAHVAYSEVDGQVFGQGRMLHV 475

## RESULT 14

A40252

elav protein - fruit fly (*Drosophila virilis*)C:Species: *Drosophila virilis*

C&gt;Date: 06-Dec-1991 #sequence\_revision 03-Apr-1992 #text\_change 02-Jul-1998

C:Accession: A40252

R:Yao, K.M.; White, K.

Mol. Cell. Biol. 11, 2994-3000, 1991

A:Title: Organizational analysis of elav gene and functional analysis of ELAV protein

A:Reference number: A40252; MUID:91246165; PMID:1903840

A:Accession: A40252

A&gt;Status: preliminary

A:Molecule type: DNA

A:Residues: 1-521 &lt;YAO&gt;

C:Genetics:

A:Gene: FlyBase:FBgn0013110

A:Cross-references: FlyBase:FBgn0013110

C:Superfamily: unassigned ribonucleoprotein repeat-containing proteins; ribonucleopro

F:16-166/Region: glutamine/alanine-rich

F:188-268/Domain: ribonucleoprotein repeat homology &lt;RRM1&gt;

F:287-355/Domain: ribonucleoprotein repeat homology &lt;RRM2&gt;

F:441-508/Domain: ribonucleoprotein repeat homology &lt;RRM3&gt;

Query Match 4.5%; Score 226.5; DB 2; Length 521;  
Best Local Similarity 19.5%; Pred. No. 8.7e-05;  
Matches 89; Conservative 102; Mismatches 154; Indels 111; Gaps 17;

QY 552 MERTEGSLEDWDIAVQKTEITRLA-----RVNEQRMKAAKEALVQOEKEAEQK 602  
DB 51 LQOQQQQVQQAIIQVQQOQTQQAIAAAAAAVTQQLQOQQQAVVAQAVVQOQQOQQOQQ 110  
QY 603 RARA--EKKALKKKKKIRGPEKGADEDEKEWGD--DEEQPSKRRRVENSIPAAGTQ 658  
DB 111 QQQVQQOQQVQQQAVVAQVQOQQOQQOQQOQQOQQOQQVQOQQOQQOQQOQQNTNG 170  
QY 659 NVEVAAGPAGCAAVDVEPPSKQKEAKASLRDMPKVLHDSKSDSITVFSNLPYSMQEP 718  
DB 171 N-----SGAQNGS-----NGTETFTNLIVNLYFQTTWTD 201  
QY 719 DTKLRPLFEACGEVVOIR-----PIFSNRGDFRGYCYVFEKEKSALQA-- 762  
DB 202 E--IRSLFSSVGEIESVKLIRDKSOVYIDPLNPOAPSKGQSLGYGFVNVVRPQDAEQAVN 259  
QY 763 ----LEMDRKSVEGRPMFVSPCVDKSKNPKVFYRSTSLKHLFTISGLPFCSTKELE 818





**THIS PAGE BLANK (USPTO)**  
**BEST AVAILABLE COPY**

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 18, 2003, 13:46:53 ; Search time 15 Seconds  
(without alignments)  
2662.780 Million cell updates/sec

Title: US-09-763-985A-2

Perfect score: 4994

Sequence: 1 MATAAETSASEPEAESKAGP.....AATEAPKMSNADFAKFLRK 963

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query %	Match	Length	DB ID	Description
1	305	6.1	650	1	NUCL_XENLA	P20397 xenopus lae
2	272	5.4	414	1	NSRL_YEAST	P27476 saccharomyc
3	268.5	5.4	702	1	CRN_DROME	P17886 drosophila
4	262	5.2	694	1	NUCL_CHICK	P15771 gallus gall
5	245.5	4.9	706	1	NUCL_HUMAN	P19338 homo sapien
6	244.5	4.9	706	1	NUCL_MOUSE	P09405 mus musculu
7	238.5	4.8	960	1	K682_HUMAN	Q9Y4C8 homo sapien
8	225	4.5	519	1	ELAV_DROVI	P23241 drosophila
9	224.5	4.5	713	1	NUCL_MESAU	P08199 mesocricetu
10	224	4.5	712	1	NUCL_RAT	P13383 rattus norv
11	220.5	4.4	1938	1	MYHD_HUMAN	Q9UKX3 homo sapien
12	217.5	4.4	1972	1	MYHB_MOUSE	O08638 mus musculu
13	216	4.3	1130	1	YL17_CAEEL	Q11102 caenorhabdi
14	211.5	4.2	500	1	GAR2_SCHPO	P41891 schizosacch
15	210	4.2	483	1	ELAV_DROME	P16914 drosophila
16	209	4.2	978	1	RA50_AQUAE	O67124 aquifex aeo
17	208.5	4.2	496	1	U2AF_CAEEL	P09078 caenorhabdi
18	208.5	4.2	1972	1	MYHB_HUMAN	P35749 homo sapien
19	207	4.1	1959	1	MYH9_CHICK	P14105 gallus gall
20	206	4.1	315	1	ROC4_NICSY	P19683 nicotiana s
21	205.5	4.1	1972	1	MYHB_RABIT	P35748 oryctolagus
22	204	4.1	475	1	U2AF_HUMAN	P26368 homo sapien
23	204	4.1	475	1	U2AF_MOUSE	P26369 mus musculu
24	203.5	4.1	1937	1	MYH8_HUMAN	P13535 homo sapien
25	202.5	4.1	1938	1	MYH8_RAT	P02563 rattus norv
26	202	4.0	733	1	SUF_DROME	P25991 drosophila
27	200.5	4.0	1939	1	MYH6_MESAU	P13539 mesocricetu
28	199	4.0	488	1	U2AF_CAEER	P09072 caenorhabdi
29	197.5	4.0	845	1	SCP1_MESAU	O60563 mesocricetu
30	197.5	4.0	1935	1	MYSS_CYPCA	Q90339 cyprinus ca
31	195.5	3.9	1935	1	MYH7_HUMAN	P12883 homo sapien
32	194.5	3.9	674	1	CMF4_SCHPO	P87312 schizosacch
33	193	3.9	1934	1	MYH7_MESAU	P13540 mesocricetu

## ALIGNMENTS

### RESULT 1

ID	NUCL_XENLA	STANDARD;	PRT;	650 AA.
AC	P20397;			
DT	01-FEB-1991 (Rel. 17, Created)			
DT	01-FEB-1994 (Rel. 28, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Nucleolin (Protein C23).			
OS	Xenopus laevis (African clawed frog).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;			
OC	Xenopodinae; Xenopus.			
OX	NCBI_TaxID=8355;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Ovary;			
RX	MEDLINE=93181171; PubMed=8441611;			
RA	Rankin M.L., Heine M.A., Xiao S., Leblanc M.D., Nelson J.W.,			
RA	Dimario P.J.;			
RT	"A complete nucleolin cDNA sequence from Xenopus laevis.";			
RL	Nucleic Acids Res. 21:169-169(1993).			
RN	[2]			
RP	SEQUENCE OF 125-650 FROM N.A.			
RX	MEDLINE=89252811; PubMed=2656405;			
RA	Caizergues-Ferrer M., Mariottini P., Curie C., Lapeyre B., Gas N.,			
RA	Amalric F., Amaldi F.;			
RT	"Nucleolin from Xenopus laevis: cDNA cloning and expression during			
RT	development.";			
RL	Genes Dev. 3:324-333(1989).			
CC	FUNCTION: NUCLEOLIN IS THE MAJOR NUCLEOLAR PROTEIN OF GROWING			
CC	EUKARYOTIC CELLS. IT IS FOUND ASSOCIATED WITH INTRANUCLEOLAR			
CC	CHROMATIN AND PRE-RIBOSOMAL PARTICLES. IT INDUCES CHROMATIN			
CC	CONDENSATION BY BINDING TO HISTONE H1. IT IS THOUGHT TO PLAY A			
CC	ROLE IN PRE-RNA TRANSCRIPTION AND RIBOSOME ASSEMBLY.			
CC	SUBCELLULAR LOCATION: Nuclear; nucleolar.			
CC	!- SIMILARITY: CONTAINS 4 RNA RECOGNITION MOTIFS (RRM).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; X63091; CAA44805.1; -			
DR	PIR; A30166; A30166.			
DR	PIR; S18874; S18874.			
DR	PIR; S30250; S30250.			
DR	HSSP; P11940; 1CWJ.			
DR	InterPro; IPR000504; RNA_rec_mot.			
DR	Pfam; PF000076; rrm; 4.			
DR	SMART; SM00360; RRM; 4.			
DR	PROSITE; PS50102; RRM; 4.			
DR	PROSITE; PS00030; RRM_RNP_1; 3.			
KW	Nuclear protein; Phosphorylation; Methylation; DNA-binding; Repeat;			

Q04836 arabidopsis  
P12036 homo sapien  
Q15149 homo sapien  
Q15431 homo sapien  
P12847 rattus norv  
Q62812 rattus norv  
P55937 mus musculu  
Q28644 spinacia ol  
Q02566 mus musculu  
P35579 homo sapien  
O95613 homo sapien  
P02564 rattus norv

34 192.5 3.9 329 1 ROC3\_ARATH  
35 191.5 3.8 1020 1 NFH\_HUMAN  
36 191 3.8 4684 1 PLE1\_HUMAN  
37 190.5 3.8 976 1 SCP1\_HUMAN  
38 189 3.8 1940 1 MYH3\_RAT  
39 189 3.8 1961 1 MYH9\_RAT  
40 188.5 3.8 1325 1 G160\_MOUSE  
41 188 3.8 233 1 ROC1\_SPIOL  
42 187.5 3.8 1938 1 MYH6\_MOUSE  
43 187.5 3.8 1960 1 MYH9\_HUMAN  
44 187.5 3.8 3321 1 PCN2\_HUMAN  
45 187 3.7 1935 1 MYH7\_RAT

KW RNA-binding. 0 BY SIMILARITY.  
FT INIT\_MET 119 134 ASP/GLU-RICH (ACIDIC).  
FT DOMAIN 155 165 ASP/GLU-RICH (ACIDIC).  
FT DOMAIN 183 202 ASP/GLU-RICH (ACIDIC).  
FT DOMAIN 232 308 RNA-BINDING (RRM) 1.  
FT DOMAIN 324 398 RNA-BINDING (RRM) 2.  
FT DOMAIN 414 487 RNA-BINDING (RRM) 3.  
FT DOMAIN 502 577 RNA-BINDING (RRM) 4.  
FT DOMAIN 583 643 ARG/GLY/PHE-RICH.  
FT MOD\_RES 154 154 PHOSPHORYLATION (BY SIMILARITY).  
FT CONFLICT 214 214 P -> Q (IN REF. 2).  
FT CONFLICT 218 219 PE -> LR (IN REF. 2).  
FT CONFLICT 410 410 E -> Q (IN REF. 2).  
FT CONFLICT 580 580 D -> E (IN REF. 2).  
SQ SEQUENCE 650 AA; 70064 MW; 5BF9BF09768E71B4 CRC64;

Query Match 6.1%; Score 305; DB 1; Length 650;  
Best Local Similarity 26.5%; Pred. No. 3e-09;  
Matches 104; Conservative 70; Mismatches 145; Indels 74; Gaps 14;

QY 552 MERTEGLEDWDIAVOKTETLARVNEORM--KAAE-----KEAALVQOEERK 597  
DB 33 MEEDSSDEVEVPVAKTPAKTATPAKATPGKAATPGKKGATPAKNGKQAKKQSEEEE 92  
QY 598 AEQRKRAAEKALKK---KKIRGPEKRGADDEDEKKGWDDDEEQ-----PS 642  
DB 93 DSDDEAEQDKPIKKNPVAKKAVAKKESEEDDDDEDE--SEEEKAVAKKTPAKKPAG 149  
QY 643 KRRRVENSPAGETQNVVAAGPAGKCAAV-----DVEPPSKQKEKAASLR 690  
DB 150 KQOESEEDDESEDEPMEVAPALGKKTQAADDEEDDEEDDDDEEDDEEQQSAGR 209  
QY 691 --DMPKVL-----HDSSKSDITVFSNLPYSMOEPDK--LRPLFEACGEVVOIRPF 739  
DB 210 KEMPTIPEAKTKTDYDASEGLSIFIGNLSKTFDELKDALREFFSKKNITQDIRI- 268  
QY 740 SNRGDFRGYCYVEFEKESALQALMDRKSVEGRPMFVSPCV--DKSNPDPKVPFRYSTS 797  
DB 269 --GNSKFGYVDFSSSEVERKALTKGKILGTEVKIEKAMAFDKNKTAENKKER--- 321  
QY 798 LEKHLFISGLFSPCTKELEICAHGTVDKLRVNTNAGPKGKGLAYVEYENESQASQA 857  
DB 322 -DSRTLFLVKNIPYSTVEELQEIFE---NAKDIRIPTGDKGNSKNGIAYVEFSNEDEANKA 377  
QY 858 VMKMDGMTIKENIKVAIS-----NPPORKVPE 885  
DB 378 LEEKQGAIEGRSIFVDFTGERSQNSGNKKGPE 410

## RESULT 2

NSR1\_YEAST  
ID NSR1\_YEAST STANDARD; PRT; 414 AA.  
AC P27476;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Nuclear localization sequence binding protein (p67).  
GN NSR1 OR YGR159C OR G7001.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-S288c;  
RX MEDLINE=91117946; PubMed=1706724;  
RA Lee W.-C., Xue Z., Melese T.;  
RT "The NSR1 gene encodes a protein that specifically binds nuclear  
localization sequences and has two RNA recognition motifs.";  
RL J. Cell Biol. 113:1-12(1991).  
RN [2]  
RP SEQUENCE FROM N.A.

RC STRAIN-S288c;  
RX MEDLINE=92355583; PubMed=1644811;  
RA Kondo K., Inouye M.;  
RT "Yeast NSR1 protein that has structural similarity to mammalian  
nucleolin is involved in pre-rRNA processing.";  
RL J. Biol. Chem. 267:16252-16258(1992).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-S288c;  
RX MEDLINE=96158062; PubMed=8585325;  
RA Skala J., Nawrocki A., Goffeau A.;  
RT "The sequence of a 27 kb segment on the right arm of chromosome VII  
from Saccharomyces cerevisiae reveals MOLL1, NAT2, RPL30B, RSRL1, CYS4,  
PEM1/CHO2, NSR1 genes and ten new open reading frames.";  
RL Yeast 11:1421-1427(1995).  
RN [4]  
RP SEQUENCE OF 188-414 FROM N.A.  
RA Rieger M., Mueller-Auer S., Brueckner M., Schaefer M.;  
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP DNA-BINDING.  
RX MEDLINE=95098604; PubMed=7800479;  
RA Lin J.-J., Zakian V.A.;  
RT "Isolation and characterization of two Saccharomyces cerevisiae genes  
that encode proteins that bind to (TGI-3)n single strand telomeric  
DNA in vitro.";  
RL Nucleic Acids Res. 22:4906-4913(1994).  
CC -|- FUNCTION: INVOLVED IN PRE-RRNA PROCESSING. SPECIFICALLY BINDS  
NUCLEAR LOCALIZATION SEQUENCES. CANDIDATE FOR A RECEPTOR AT THE  
NUCLEUS THAT MAY BE INVOLVED IN BOTH RNA AND PROTEIN TRANSPORT.  
CC BINDS TELOMERIC SEQUENCES OF THE TYPE (TG[1-3])N IN VITRO.  
CC -|- SUBCELLULAR LOCATION: NUCLEAR; POSSIBLY AT THE NUCLEOLUS.  
CC -|- INDUCTION: IN RESPONSE TO LOW TEMPERATURE (BY COLD-SHOCK).  
CC -|- SIMILARITY: BELONGS TO THE GAR FAMILY.  
CC -|- SIMILARITY: CONTAINS 2 RNA RECOGNITION MOTIFS (RRM).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See http://www.isb-sib.ch/announce/  
or send an email to license@isb-sib.ch).  
CC -----  
DR EMBL; X57185; CAA40472.1; -;  
DR EMBL; X85807; CAA59817.1; -;  
DR EMBL; 272944; CAA97173.1; -;  
DR EMBL; 272946; CAA97180.1; -;  
DR PIR; A39205; A39205.  
DR SGD; S0003391; NSR1.  
DR InterPro; IPR00304; RNA\_rec\_mot.  
DR Pfam; PF00076; rrm; 2.  
DR SMART; SM00360; RRM; 2.  
DR PROSITE; PS50102; RRM; 2.  
DR PROSITE; PS00030; RRM\_RNP\_1; 2.  
KW DNA-binding; RNA-binding; Nuclear protein; Repeat; rRNA processing.  
FT DOMAIN 29 136 SER/ASP/GLU-RICH.  
FT DOMAIN 168 246 RNA-BINDING (RRM) 1.  
FT DOMAIN 267 345 RNA-BINDING (RRM) 2.  
FT DOMAIN 366 384 RNA-BINDING RGG-BOX (BY SIMILARITY).  
SQ SEQUENCE 414 AA; 44535 MW; 90DEEE7B8C20BC0C CRC64;

Query Match 5.4%; Score 272; DB 1; Length 414;  
Best Local Similarity 22.2%; Pred. No. 9.5e-08;  
Matches 84; Conservative 84; Mismatches 172; Indels 38; Gaps 9;

QY 554 RTEGSLDWDIAVOKTETLARVNEORMKAAEKAALVQOEERKAR----- 605  
DB 6 KVGKNGKVKASKQAKKEKAKAVSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 65  
QY 606 -----AEKKALKKKKIRGPEKRGADDEDEKKGWDDDEEQPSKRRRVENSIPAAGET 657  
DB 66 SSSSSSSSSSEAEETKKEESKSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 121















[illegible]





```
Db 1596 RQLHEYTELEDERKORALAAAANKKLEGLDKOLEQADSAIKGREBAIKQLRKLOQMK 1655
QY 623 RGADDEDEKEWGDDE-----EQPSKRRRRVENSIPAAGETQNVVEAAGPAGKCAAVDVEP 677
Db 1656 DFQRELDARASRDEIFATSKENAKSLKLEADLMQLE-----DLAAAEARAKQA-DLEK 1710
QY 678 PSQKQKAAKSLKMDPKVLHDSKDSITVFNPLPSMOBPDTKLRLPLFACGEVVOIRP 737
Db 1711 EELAEELASSLSG--RNTLQD-EKRRLEARIAQLEEEBEEQCNMEASDVRVKATLQAE 1767
QY 738 IFNRGDFRCYCYVEKKEKSAQALEMDKSVSGRPMFVSPCDVSKNPDKVFVRYSTIS 797
Db 1768 OLSN-----ELATERSTAKKNESARQQLERO-----1793
QY 798 LEKHLKLFISGLPFSCTEKEEETCKAHGTVKDLRLVTNRAGPKGLAYVEYE-----NES 852
Db 1794 -----NKELRSKLQVEGAVK-AKLASTVALEAKIAQLEEQVEQAREK 1837
QY 853 QASQAVKMDGWTIKENITKVAISNPPQKVPKPKTRKAPGPMILLPOTYARGKGRTO 912
Db 1838 QAATKSLKOKKKLKEVLLQV-----EDERKMAEQYKEQAE-----KGNTK 1878
QY 913 LSLPLRALQ 921
Db 1879 VKQLRKOLE 1887

RESULT 13
YL17_CAEEL
ID YL17_CAEEL STANDARD; PRT: 1130 AA.
AC Q1102;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 131.5 kDa protein C02F12.7 in chromosome X.
OS C02F12.7.
GN Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fristol N2;
RA Miller N.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: WEAK, TO MYOSINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U41545; AAK39135.1; -
DR WormPep; C02F12.7; CE03901.
KW Hypothetical protein; Coiled coil.
FT DOMAIN 121 779 COILED COIL (POTENTIAL).
FT DOMAIN 805 1061 COILED COIL (POTENTIAL).
SQ SEQUENCE 1130 AA; 131485 MW; B0FD2FE3D99FB09 CRC64;

Query Match
Best Local Similarity 22.0%; Pred. No. 0.00035;
Matches 172; Conservative 123; Mismatches 261; Indels 226; Gaps 38;

QY 174 EKAVDYICPNLWIEYGVSGIGQKGLKVRFSRALS SVGLHMTKGLALWE-----229
Db 135 EETVAEY-----ERQKYNVMTFS-----EYRERVAERERKLEA EYSKIIALSEEVLG 183
QY 230 AYREFES-----AIVEAARLEKHSL--FRQLAIPLYDMEATFAEYEEWSEDPPEV 281
Db 184 AKKDFEARKMSFOALQDFEREKEQALEKLRKEHQEQVLEQRFSDTQLNLNLE-----237
```

## RESULT 14

## GAR2\_SCHPO

```
ID GAR2_SCHPO STANDARD; PRT: 500 AA.
AC P41891; O13707;
DT 01-NOV-1995 (Rel. 32, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Protein gar2.
GN GAR2 OR SPAC140.02.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=95319932; PubMed=7596817;
RA Gulli M.-P., Girard J.-P., Zabetakis D., Lapeyre B., Melese T.,
RA Caizergues-Ferrer M.;
RT "gar2 is a nucleolar protein from Schizosaccharomycetes pombe required
RT for 18S rRNA and 40S ribosomal subunit accumulation.";
RL Nucleic Acids Res. 23:1912-1918(1995).
```



